

***Final Exit*: euthanasia guide sells out**

A Hemlock Society book on suicide is leading the best-seller list in the 'how to' category, revealing widespread public concern about having control at the end of one's life.

CONTRARY to the admonition of the poet to rail against the dying of the light, most people have a deep desire to go gently into that good night when the time comes. If the medical profession needed proof of people's desire to have some control over their demise, it need only heed the fact that a slim book called *Final Exit* has hit the best-seller list in the United States where suicide is still taboo but fear of twilight life tethered to feeding tubes or respirators is widespread.

Written by Derek Humphry, a British journalist and euthanasia proponent, and president of the Hemlock Society, which has headquarters in the state of Oregon, *Final Exit* describes what the author calls the "practicalities of self-deliverance". In considerable detail, Humphry reviews a variety of options for suicide for the dying and presents a chart giving combinations and dosages of lethal drugs.

It is not surprising but nonetheless disappointing that *Final Exit* has come under attack by ethicists and others who piously fear that its prescriptions for suicide for the terminally ill will be abused by the merely downhearted. Thus newspaper accounts of the book quote 'experts' worried that it will get into the hands of the mentally ill without offering them the opportunity for what is now a favourite US pastime — 'counseling'. To worry about such things first suggests that despondent people cannot think up ways of committing suicide without the aid of a how-to guide and, second, indicates that the worriers have not read Humphry's book.

Final Exit is about release from the pain of terminal illness, not about the romance of suicide. There are no consumptive young maidens dying gracefully on Victorian couches. "Roughly half the people who die in Western society currently are connected to equipment", Humphry claims. In theory, they have the option of pulling the plug. The suicide guide is intended for the dying for whom there is no plug to pull, and only for those who have already decided that self-deliverance is ethically acceptable. "If you consider God the master of your fate, then read no further", Humphry says: but if you and your loved ones want to have control over the time of death, read on and plan ahead. Gentle suicide is not easy.

Shooting oneself in the head is not only violent and bloody, it is often not successful. Dying in a closed garage with the motor running takes long enough that

discovery is likely. Ingesting drugs on a full stomach leads to vomiting, not dying. And cyanide can lead to a painful end, as witnessed by the deaths from cyanide-laced Kool-Aid a few years ago in Jonestown.

The gentlest and surest exit for the terminally ill — often individuals too incapacitated to commit suicide unassisted — lies with help from the healing profession which, in the final analysis, is what the Hemlock Society is all about. The barbiturates Seconal and Nembutal are recommended for those who can still swallow. A variety of injectable drugs is offered for the comatose.

Final Exit is the first of several Hemlock Society books to be distributed by a commercial publisher — a small house called Carol Publishing in Secaucus, New Jersey. The tale of the book's success is the tale of public appetite for information on this hidden subject. Between April and mid-July, 2,000 copies out of a print run of 8,000 were sold. A story in the *Wall Street Journal* stimulated further publicity. In Washington, DC, bookstores have customers on waiting lists. Total sales of more than 100,000 are anticipated.

Euthanasia is a subject ripe for wide discussion in nations where physicians armed with the life-sustaining tools of modern science may be violating their pledge to 'first do no harm' in their zeal to ward off death. Among Western nations, only the Netherlands has an unambiguous policy that allows physicians to ease the passage of the dying. The doctor's role was highlighted recently when a physician in Rochester, New York, wrote a journal article describing how he assisted in the death of a leukaemia victim by prescribing lethal drugs for her to take when she was ready. That the physician was brought before a grand jury reveals the state of the law. That the grand jury refused to indict reveals something about the public's attitude. Unfortunately, it cannot yet be read as the voice of the majority.

But the realities of *Final Exit* can no longer be hidden. As the average age of the population increases, the abstract debates of ethicists about informed consent and of physicians about the duty to heal will come face to face with realities and indignities of incapacitating illness that in an earlier age would have claimed its victims swiftly. The 'ethic' of using technology to prolong life when life is ebbing should give way to the ethic of individual choice, including the right of physicians to choose to assist the dying. □



Past imperfect

Publications more than 25 years old are likely to be forgotten, which is a shameful waste.

THE letter from Professor Edward Harrison on page 574 of this issue should not be mistaken for yet another claim to priority for discovery (although it is that). The letter is more important for what it implies about the way in which even the relatively recent literature can quickly fail to serve as the vivid record of past discovery that it is generally supposed to be. The point at issue is the resolution of Olbers' paradox — the remark early in the nineteenth century that, if the number of stars in the Universe were so large as to be an approximation to infinity, the whole surface of even the night sky should be as bright as the surface of the Sun (or whatever is the brightness of a typical star).

Early in the nineteenth century, it had become clear that the number of observable stars increased with each improvement of technique, suggesting that the number might not be very different from infinity, but the notion that the Universe might be expanding, and that light from distant stars might on that account be shifted towards the less bright red end of the spectrum, had not yet arisen. So Olbers' paradox could be used in two ways — either to assert that the number of stars in the Universe (other galaxies had not then been discovered) could not, after all, be infinite or anything like it, or to focus attention on the need to explain why the night sky is not uniformly bright.

The doctrine of the expanding Universe radically changed the terms of reference in two ways. First, the brightness of receding stars will be diminished by the recession; second, an expanding universe must expand from something and so must have a finite age, raising the question whether there has been enough time since the beginning of the Universe for there to have been created enough stars (or galaxies) for the premise of Olbers' paradox (that the number of stars or galaxies is essentially infinite) to be satisfied. So the resolution of the paradox hinges on the relative importance of recession and finite age.

What Harrison now says is that, more than a quarter of a century ago, he had argued in a contribution published in this journal that the effect of recession is much less important than that of the finite age of the Universe. That is also the conclusion of an argument published earlier this year by Paul S. Wesson, which attracted an approving comment in *Nature*; neither Wesson's article nor the account of it in this journal referred to Harrison's much earlier work.

The tone of Harrison's moderate letter prompts one interesting conclusion — that even he is not surprised that a piece of work published in 1964 should have been forgotten by 1991. (It is more shocking that this journal's commentator should have failed to draw attention to

Harrison than that Wesson should not have done so.) But sadly, that must these days be a common experience. There is a sense in which the literature is rendered inaccessible by the mere passage of time. Of course, outstanding discoveries make their way into the textbooks and monographs from which young people learn science, but the likelihood that other material will be retrievable necessarily depends on the idiosyncracies of individual memories.

There is no early prospect that matters will improve. Although machine retrieval could in principle present interested readers with references to everything previously published in some nominated field, it is unthinkable that such a service would be feasible for the years before the creation of the first machine databases (in the early 1970s). And even then there are obvious intellectual difficulties, notably those of constructing cumulative indices that will retain their meaning as the decades pass. So it may be necessary that we should become reconciled to the inadvertent neglect of earlier work, despite the time and talent wasted in its duplication. □

Circles within circles

Are the mysterious circles of flattened corn in British cornfields simply a media ploy?

IF ultimately of no other scientific interest, the annual spate of crop circles in British fields brings to light the circular nature of science itself — at least as far as the media are concerned. Stories have appeared over the summer in the British press declaring the file on the inscrutable circles closed at the hands of some expert or other. Strangely enough, however, the same theories had all been advanced at this time last year.

Most entertaining, and hence most appealing to a desperate editor, is the theory that the rings of flattened grain are carved by birds chasing each other round in circles as part of some seasonal mating ritual. No matter that the circles are all far too perfect, and that the process has never been observed. Animals are funny; sex is funny; run the story each year as a 'discovery'.

The same entertainment value obviously does not adhere to the proposition that the circles are some kind of electrostatic phenomenon brought about by an unknown, yet firmly prosaic, combination of climatic and geographical factors. It was a great relief to everybody, then, when it emerged that this year the boring explanation was to be investigated by a team of scientists armed with expensive photographic equipment and characteristically taking something too seriously. No surprise then that the electrostatic explanation cropped up again this summer.

Whatever the eventual outcome of the crop circle saga, it is unlikely to stand out from the perennial volley of recycled 'discoveries' in the British press, which shows no signs of abating in the years to come. □

French scientists rebel against review system

- Lawsuits charge favouritism, vote trading
- New rules passed to improve process

Paris

WIELDING the power of the courts, mass petitions and the scientific trade unions, researchers in the laboratories of the top French science agency are rebelling against what they claim is a scientific review process that has become corrupt. Government documents obtained last week reveal that about 60 French scientists are now engaged in lawsuits against the Centre National de la Recherche Scientifique (CNRS), some charging favouritism, illegal lobbying, and vote trading in the review committees that determine the future of more than 22,000 government researchers.

Although government officials say that approximately two thirds of the cases are based on such individual issues as radiation exposure and pension eligibility, at least 20 lawsuits stem from researchers who claim to have been victims of bias at the hands of the review committees. According to government records, a total of 28 cases are at the first court level (known as an administrative tribunal), three are in the court of appeals, and 28, including the 20 dealing with review issues, are in the highest court.

In a sign of growing recognition at CNRS that it has a problem on its hands, Claude Paoletti, director of the agency's life sciences programme, sent a memorandum in May to all CNRS directors warning there should be no more "informal communication" with scientists who have brought suit against the agency.

"CNRS has paid special attention to reducing the constraints and easing the procedures which could hinder [a free scientific] dialogue", Paoletti wrote. "It is in this climate of mutual confidence that research is, as one might say, 'administered'. However, over the past few years, persons who are nevertheless members of our community have chosen to depart from this confidence and serenity and have gone to trial to settle cases in which they are in opposition to the direction of CNRS." Paoletti cited "about 100" such cases at the time.

Jean Marie Bertrand, the CNRS secretary general, says he does not find the number of cases especially alarming. "A government organization like CNRS has to use administrative procedures to settle disputes," he says. "Many kinds of problems that do not require judicial procedures in the United States do in our system. Sixty [the official total of current cases] is not such a big number when you

have 25,000 employees [including support staff]."

But researchers say that conflict with management over scientific review has become a real problem in recent years. And when asked for a culprit, many point to arcane bureaucratic rules that have allowed scientific review groups to become virtual dynasties.

France's national scientific review committees are split between members elected by the scientists and those appointed by the science agencies themselves. CNRS has historically interpreted French law as allowing committee members to be reappointed indefinitely, so some scientists have served on essentially the same committee — although its name may have changed — for several decades. Adding to that is what appears to many as conflict of interest: in addition to the research of all CNRS scientists, the committees are also allowed essentially to review themselves.

"The tradition in France has been that the members of an evaluation committee are themselves evaluated — with great indulgence — together with their team by the committee to which they belong", says Jacques Ninio, a biologist at the CNRS Institut Jaques Monod who sued CNRS in 1990 after he was denied promotion to the equivalent of full professor for the seventh consecutive year.

As public servants, Ninio says, "we are subject to such a heap of constraining legal and administrative rules that nothing can be done without a certain amount of cheating". If a laboratory director wants to offer a permanent position to a technician who has worked for years within the laboratory on grants alone, he must first issue a public job offer, to which as many as 40 unemployed candidates may apply. "Every one will go through an exhausting selection procedure before an 'objective' audit committee, usually ending with the predictable victory of the insider", he says.

A survey conducted in April by the French scientific trade union SNCS at INSERM, the biomedical research agency which has a review committee structure like CNRS, found that applicants for government research positions who had a close colleague on the review committee had 4.6 times the chance of being selected as those who did not.

Yet despite the obvious biases in the selection process, researchers might not have to resort to the courts if scientific

AIDS

Leaked report queries Gallo patent

Washington

A DOCUMENT purporting to be an official government analysis of the validity of the AIDS patent awarded to scientists at the US National Institutes of Health in 1984 has been leaked to the *Chicago Tribune*, which ran a story headed "Probe finds fraud in AIDS studies" in its 11 August issue. According to the *Tribune*, the document says that AIDS researcher Robert C. Gallo made "untrue and misleading statements" in a sworn statement about the development of a blood test for AIDS. The *Tribune* adds that "the report offers no evidence to show that Gallo knew or should have known that some of the statements were untrue".

Attorneys for Gallo and his colleague Mikulas Popovic are livid not only about the leak but also about the report itself, which they find "nonsensical" and "full of errors". The report is presumed to have been written by Suzanne Hadley, who has recently been pulled off the case by NIH director Bernadine Healy.

One issue, for instance, is whether Gallo told the truth when he said he was growing virus in 'continuous' culture for five months when, as stated in the *Science* paper, the culture was re-fed with fresh cells along the way. The report finds the use of the word 'continuous' to be untruthful, according to the *Tribune*. Whether anything in this leaked document has real bearing on the validity of the US AIDS patent is not clear at present.

Barbara J. Culliton

trade unions like SNCS had more clout with the government. In the last six years the unions have lost over 50 per cent of their membership — a reflection, researchers say, of a decreasing confidence in their effectiveness and independence.

The picture, however, may be improving. Earlier this year, CNRS instituted several new rules that it hopes will improve the process and quell some of the criticism. Perhaps the most significant reform is that the agency will no longer let committee members serve more than two four-year terms, something that officials hope will bring in new blood. CNRS also announced that it would reduce the number of national scientific review committees from 49 to 40. "This is a way to reduce lobbying and 'unclear attitudes' among the committee members, by bringing together more people from different communities" who can police each other, says Bertrand. Despite the problems, he adds, "it is a great advantage to have scientists reviewing science. We want to keep that." In short, CNRS is counting on the reforms to win the day before the lawsuits do.

Christopher Anderson

Exploring the limits to continued growth

Washington

WITH both Congress and the Bush Administration determined to limit spending on the indirect costs of research, and growth in the National Institutes of Health (NIH) budget constrained under the budget agreement between the two arms of federal government, 1991 might not seem the best time for US biomedical research centres to expand. Yet expansion is exactly what many institutes across the United States are planning, to the alarm of several Washington-based biomedical research policy groups.

Federal government spending on biomedical research has grown steadily over the past decade, from less than \$5,000 million in 1980 to today's annual budget of around \$10,000 million. But the harsh financial reality of the US budget deficit means that this growth is now likely to slow, at least temporarily. Ed Stamm, vice president of the Association of American Medical Colleges (AAMC), is concerned that the expansion plans of his member colleges will, in total, greatly

exceed the resources available for biomedical research over the coming few years. "I hate to make the comparison," he says, "but how many McDonalds can you build? There's probably some limitation on the hamburger market." AMC has already launched a pilot study to assess just how over-ambitious the plans of its members are.

The Association of Academic Health Centers (AAHC) is similarly concerned. The priority for its science policy taskforce over the next 12 months is to examine how AAHC members can raise money for capital investment in new buildings and equipment in the face of the current financial difficulties. AAHC president Roger Bulger is especially worried about the fallout from the current indirect research costs scandal on investment in medical schools in the state university system. Throughout the 1960s and the 1970s, he notes, state legislatures made the expansion of their medical schools a matter of public policy. But with the current scrutiny of indirect costs

revealing the extent of federal support for capital investment in state schools, says Bulger, state legislatures are saying: "Why should we do it?"

The arguments of AAHC and AAMC assume that the predicted shortfall in investment in new biomedical research facilities and in research grant support will not be made up by private foundations and by the pharmaceutical industry, as has happened in Britain to a certain extent over the past decade. But that seems a reasonable supposition. The only significant charitable player in the US biomedical research funding field, the Howard Hughes Medical Institute, now plans to keep its spending at a steady level, after five years of growth. And David Korn, dean of Stanford Medical School, says he does not see industry "champing at the bit to join the fray".

The concern of the Washington-based umbrella organizations is noted, but swiftly dismissed by most of their member schools. Some medical school

Philadelphia: US biomedicine in microcosm

Jefferson Medical College epitomizes the research expansion witnessed at many US medical schools over the past decade. Languishing in the lower ranks of US biomedical research centres in the late 1970s, the school has pulled itself up by its bootstraps and now boasts a healthy reputation for research.

Yet a broader look at Jefferson's home city, Philadelphia, reveals the difficulties that may face medical research centres that are less successful in the race to expand. Jefferson's latest coup — the recruitment of cancer geneticist Carlo Croce, pharmacologist Gerald Litwack and an entourage of some 200 principal investigators, postdocs, graduate students, technicians and administrators — was won at the expense of neighbouring Temple University's Fels Institute. Robbed of its best researchers, and with its host university in financial crisis, Fels faces an uncertain future.

Jefferson's rise to prominence was engineered in large part by its charismatic dean, Joseph Gonnella. An Italian native, Gonnella embarked on an aggressive policy of expansion and acquisition in research from the early 1980s. The first major victory came in 1985, with the recruitment of Darwin Prockop, the collagen biochemist, from Rutgers. Others soon followed. Having consolidated Jefferson's research into connective tissue, Gonnella moved on to pathology, luring Emanuel Rubin and his research team from Hahnemann University, also in Philadelphia.

The latest wave of expansion was



Jefferson's new \$80 million building

made possible only by the construction of a new \$80 million building. The defection of Croce and his colleagues from Temple's Fels Institute to Jefferson has caused bitter recrimination between the two institutions, but Croce's decision to move is perhaps understandable. In sharp contrast to Jefferson's gleaming new facilities and relative opulence, Fels is situated in a run-down area of Philadelphia.

Allen Myers, dean of Temple's medical school, does not believe that Croce's

departure will be a mortal blow for the Fels Institute. "We're not standing still," he says, adding that a new director should be appointed in the autumn. Temple's administration is reported to be prepared to pay for the rescue of Fels, but other medical researchers in Philadelphia believe Myers faces a difficult task. Croce says that similar promises of investment were made to him when he was lured to Fels from the Wistar Institute several years ago, but that Temple never did as much as had been agreed.

Gonnella attributes Jefferson's success to a focused policy of expansion — concentrating on defined areas of research — and to the medical college's strength in providing clinical care for the residents of Philadelphia. Jefferson's finances were boosted during the 1980s by a change in the Medicare payment system. Until the early 1980s, hospitals were reimbursed for the full cost of treating Medicare patients. But under the 'prospective payment' system, a set fee is paid for each treatment. The result is that hospitals, such as Jefferson, that treat patients more cheaply and efficiently can pocket the change. In Jefferson's case, these profits have been ploughed into research infrastructure.

So is Gonnella trimming back his expansion plans in the light of the current limits to the US biomedical research budget? "The 1990s will be more difficult", he concedes, but Gonnella believes Jefferson's rise will be inexorable. Already, he has targeted neurosciences as the next area for conquest. **P.A.**

NEWS IN BRIEF

deans point out that they simply have to expand their research facilities to provide adequate laboratory space for their existing researchers, whose work is now hampered by overcrowding. "Our biggest problem is space," says James Pittman, dean of medicine at the University of Alabama at Birmingham, where some \$90 million has been invested in new biomedical research facilities over the past decade. "There are horrible fights. We see no way to slow down."

Kern Wildenthal, president of the Southwestern Medical Center at the University of Texas, where annual external research funding grew fivefold during the 1980s, says that "it's unwise for any university to expand on the assumption that quality will follow". But Southwestern urgently needs more than 30,000 square metres of additional research space, he says.

Nevertheless, some biomedical research centres are expanding even though they have no immediate need for any extra research space. The growth of Jefferson Medical College in Philadelphia (see panel) is one example, but it is by no means exceptional. "We will be recruiting new faculty and developing new programmes," says Arthur Asbury, vice dean for research at the University of Pennsylvania Medical School.

Why are some biomedical research centres behaving in this seemingly rash manner? Alabama's Pittman believes he has the answer: "I think you need a critical mass to do research. Smaller schools will have a difficult time competing." Even if the University of Alabama at Birmingham did not need to expand to ease its overcrowding problem, the expansion would continue, he suggests. "I don't think we would willingly become one of the have-nots." Penn's Asbury agrees: "Less developed institutions tend to wither in time of drought."

So in the current harsh financial environment, it seems that the US biomedical research enterprise may be destined to undergo its own version of evolution by natural selection, with the weakest schools going to the wall. If that is the case, then quality US biomedical research will, over the next few years, become increasingly concentrated into a small number of centres of excellence. That in itself may be no bad thing, but AAHC's Bulger is concerned that the transitional period would see an excessive, and wasteful, movement of researchers among institutions.

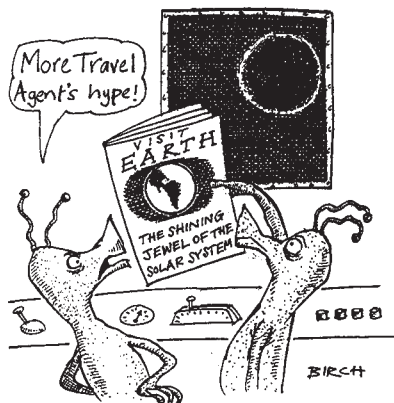
If the evolutionary analogy holds true, however, the patrician voices of AAMC and AAHC are unlikely to be heeded: no darwinian theorist would predict that any individual research centre should jeopardize its survival by forgoing its own expansion to serve the common good.

Peter Aldhous

Hazy hazy days

Washington

ASTRONAUTS aboard the US space shuttle Atlantis report that the Earth's atmosphere seems much hazier at present than on their previous trips into orbit. In a televised press conference last week, mission specialist Shannon Lucid and flight commander John Blaha speculated that the oil well fires in Kuwait or the recent volcanic eruptions in the Pacific might be



to blame. But according to Vic Whitehead, a geoscientist at the National Aeronautics and Space Administration's Johnson Space Flight Center, a more likely cause is a recent spate of dust storms in Africa.

The current Atlantis mission includes a survey of stratospheric ozone, with a device called the shuttle solar backscatter ultraviolet instrument. These data will be used to calibrate the ozone mapping instruments on orbiting satellites, but should also help to reveal any ozone depletion caused by the Kuwaiti fires or recent volcanic activity.

P.A.

Birth rate shows signs of decline

Washington

THE birth rate in developing countries has slowed over the past decade, according to results described at a conference in Washington last week. Martin Vaessen, head of the Demographic and Health Surveys (DHS) project, a huge study encompassing some 25 developing countries and funded by the US Agency for International Development, says the results indicate that the population control message is beginning to be heard. The most encouraging results from the conference, called to review the past five years' work of DHS, is the finding that the birth rate in many African countries has fallen over the past decade, Vaessen says. In Kenya, for example, the average number of children born to each woman has fallen from 8.2 to 6.5 over the past 12 years. Although the birth rate in Latin America has been falling since the late 1960s, the African trend is more recent.

But Vaessen warns that much remains to be done, with no signs of a declining birth rate in some other African states, such as Mali, Liberia and Senegal. And even in developing countries where the birth rate is falling, reduced mortality means population growth is still a problem.

Less animal testing

London

BRITISH companies slashed their use of animals for testing cosmetics and household products by two-thirds last year, part of an increasing trend toward alternatives to animal testing of consumer goods, according to a new government report on UK research animal use. The figures are the latest sign of a decrease in the number of procedures carried out in the United Kingdom on living animals, a total that fell from 3.32 million to 3.21 million last year, part of a 15-year decline from a peak of nearly 6 million in the mid-1970s.

One of the few areas in which animal research is increasing is in medical research on non-human primates, which climbed 13 per cent from last year to a total of 704 experiments. Although the Royal Society for the Prevention of Cruelty to Animals commended the reduction in total number of scientific procedures, it lumped it with another 'positive' trend that UK scientists may not feel so good about: a decrease in the number of British research establishments, from 375 to 366.

C.A.

Additional sums

London

BRITAIN'S principal research council intends to request a funding increase for its mathematics division after an advisory panel found that the UK mathematics funding is "dangerously inadequate" and that the country is missing opportunities to capitalize on recent advances in the discipline.

Sir Mark Richmond, chairman of the Science and Engineering Research Council (SERC), announced last week that, as recommended by the panel, the council would request a 50 per cent increase in the number of mathematics research student scholarships and a tripling of postdoctoral fellowships in the agency's bid for future funds.

Although that request is subject to government approval, SERC intends to take matters into its own hands by allocating more student fellowships to mathematics next year, although the exact number has not yet been announced. In October, the council also plans to release a request for new proposals in applied solutions to non-linear problems, part of the non-linear research initiative started last year. Total funding for that effort has not been set, but officials are hoping for as much as £1 million.

C.A.

Erasing high-tech barriers

New Delhi

INDIAN industrial research laboratories accustomed to developing technologies for a protected market are jittery following the announcement of a sweeping new Indian industrial policy, which opens the domestic market to competition, loosens official controls and gives an open invitation for investment and technologies from overseas companies and multinationals. The laboratories must now compete in order to survive.

The policy, part of a package of economic measures to rescue the crisis-ridden economy, is a departure from socialist protectionism and the tenets of technological self-reliance laid down by successive governments since India's independence in 1947. In effect, the government announcement bids farewell to earlier policies anchored on the principle of maximum support to indigenous research and development and 'selective' import of foreign technologies.

The reforms allow foreign companies to hold ownership interests of up to 51 per cent in several high-technology industries. Indian companies can now enter into technology collaboration agreements and hire foreign technicians without government clearance. Industrial licensing has been abolished for all projects except a short list of industries considered to be either of strategic value or socially or environmentally sensitive.

By dropping the fences, the government hopes to attract billions of dollars of direct foreign investment. The technologies brought by the investors will expand the production base and increase exports, thereby stimulating local industries. Many economists view the reform as a drastic but necessary step to integrate into the global economy, but it has brought India's industrial research laboratories to a crossroads.

"The new policy spells disaster to indigenous technologies in the pipeline", said an official of the National Research Development Corporation (NRDC), which is responsible for commercialization of indigenous technologies. In the past ten years, NRDC sold over 360 processes to industry. "But once the invasion [of foreign technologies] begins, we may have to close the shop", the official said. Competition is acceptable between equals, he said, but "local technologies will not even have an opportunity to stand up before those of giants like Du Pont, Dow Chemicals or Hitachi". Du Pont has announced plans to invest \$800 million in new projects in India, and Japan intends to invest \$2,000 million.

The Department of Scientific and Industrial Research, which is responsible for promoting research in industry, is also

worried. Over the years, the department has persuaded some 650 companies to establish in-house research and development units by giving them financial incentives, such as income-tax exemption. In the changed climate, the department fears the companies may decide to import the technologies instead of developing them in-house.

One of the agencies where the impact of liberalization will be seriously felt is the Council of Scientific and Industrial Research (CSIR), which presides over some 40 laboratories specifically set up to develop technologies for Indian industry. "We were prepared for a gradual reform, but it came all of a sudden", said S. K. BRAZIL

Joshi, director general of the council. CSIR, which derives 30 per cent of its budget from the sale of its technologies to industry, must now compete to survive. According to Joshi, CSIR is not threatened by competition in such areas as drugs and agrichemicals, "but we cannot compete with multinationals in areas like materials and polymers". Joshi says CSIR has no option but to reassign priorities and even abandon projects that have little chance of competing.

Joshi does not expect the impact of the new policy to be totally negative. "The fear of losing in the race will make our scientists do better", he says. Yet CSIR is hoping to seek from ministries fiscal incentives for industries that adopt CSIR technologies rather than imported ones.

K. S. Jayaraman

Opening up to computers

São Paulo

BRAZIL is getting close to abolishing its 'market reserve' on computers — the protectionist import ban that, critics say, has made local industries lag behind the rest of the world. A recently passed congressional bill will — if approved by the Senate and by president Fernando Collor de Mello — open up the Brazilian market after October 1992.

Collor has made it a priority to liberalize the country's economy, where the state's hand is seen almost everywhere. And nowhere has state control been more cumbersome than in the computer industry, where Brazilian microcomputers cost on average three times more than US or European equivalents.

The move to open Brazil to foreign competition has elicited very different reactions from the country's scientists. As might be expected, employees of Brazil's computer industry are among the most vocal defenders of the market reserve.

The Brazilian Society for the Progress of Science, for instance, argues that the import restrictions have created "local technologies" and provided jobs for thousands, including engineers and other electronics researchers. The society's president, Ennio Candotti, routinely denounces "American pressures" on the computer industry and other nationalist causes, such as pharmaceutical and biotechnology products patents.

But scientists who need computers and other electronic equipment for their work have other opinions. Biomedical researchers are particularly unhappy with the present situation. Not only do they not have access to good computers, but they must also struggle with the government bureaucracy whenever they need equipment that has electronic circuitry.

The usual way to circumvent the infor-

matics law is to resort to smuggling, and it has been said that smugglers are the real heroes of Brazilian research. There are smugglers who specialize in dealing with scientists, and some have their clientele concentrated in a single university. Estimates are difficult to make, but some say that several thousand computers are smuggled into the country each year.

Brazil's secretary of science and technology, José Goldemberg, is a strong defender of opening the market to competition. He has compromised a bit by offering to allow imports gradually, but some articles that used to be forbidden for import are now beginning to show up in local shops. When he served previously as rector of the University of São Paulo, he angered many on the staff by insisting that their work be evaluated, and now he says he wants to apply the same high standards to computer manufacture.

No one denies that opening up the market will mean that some local companies will be overwhelmed. The US chain Radio Shack is planning to open an outlet in Rio de Janeiro soon, and others will follow. One option for survival for Brazilian companies is to create joint ventures with foreign concerns. IBM already plans one such venture with a São Paulo company, SID Informática. And the Brazilian legislature is now discussing other ways to help Brazilian companies, such as tax incentives.

Even if the final legislation does not open the market entirely, it will go far in improving things for consumers — the most conspicuously absent voice in the debate. Already the timid measures taken so far date have had consequences. In an inflation-prone country where falling prices are virtually unheard of, prices for microcomputers have started to fall.

Ricardo Bonalume

SLAC accounting for damage?

Washington

STANFORD University is once again struggling to clear its name against the accusation that it has defrauded the US federal government. The latest accuser is Tim Axe, a former employee at the Stanford Linear Accelerator Center (SLAC), who alleges that his supervisor, Bill Voreyer, encouraged him to submit a fraudulently inflated estimate to the Department of Energy for the cost of repairing damage at SLAC caused by the Loma Prieta earthquake in October 1989. Axe first contacted Stanford president Donald Kennedy with his allegations last February, but when Stanford completed its own internal audit report on the incident and found no fraud, Axe approached a local California newspaper, the *San Jose Mercury News*, with his story.

In February 1990, Axe alleges, he was asked to make an estimate of earthquake damages that would support Voreyer's initial assessment of \$480,000, made in the week following the earthquake. Axe says that instead he calculated the total damage caused by the earthquake to be about \$50,000. Axe claims that Voreyer

told him, "if you find some work that needs to be done and you can't say it was done by the earthquake, go ahead and charge it to this [the earthquake damage] account". Axe also alleges that he was dismissed because of his refusal to increase his estimate. SLAC director Burton Richter says that Axe's dismissal was entirely unrelated to the earthquake damage estimate.

According to Stanford's internal audit, there are "no grounds for considering a misappropriation of funds", because the Department of Energy never provided SLAC with a specific sum of money earmarked for the earthquake repairs — instead, the laboratory received a larger-than-usual operating budget from the energy department for 1990, which SLAC could spend as it liked.

Nevertheless, Stanford officials have not yet dismissed Axe's fundamental complaint that he was asked to 'pad' his estimate to win more funds. Stanford's outside auditing firm, Coopers and Lybrand, has now been asked to examine Axe's claim that his superiors acted improperly.

Peter Aldhous

SCIENCE MAGAZINES

Discover suspends activities

Washington

Discover, the only popular US science magazine aimed at lay readers, was closed last week. The magazine's demise came as a shock to most of its editorial staff, who learned of the winding up only on 7 August — a day before it was announced publicly. Although *Discover* itself has shown a profit, its parent company, Family Media of New York, decided to shut down because of falling

Denise Graveline, AAAS head of communications, says there are no plans to acquire *Discover*. A spokesman for the National Academy of Sciences says it is unlikely that the academy will buy it.

Family Media expanded its range of titles during the 1980s, borrowing heavily to do so. Industry observers attribute the company's failure to over-expansion, in the face of the advertising recession.

The disappearance of *Discover* leaves the United States without a mainstream monthly science magazine aimed at a lay readership. (*Omni* is considered by many to have moved towards the coverage of 'fringe' science, with articles on such topics as extrasensory perception.) Two other popular science titles of the mid-1980s, *Science Digest* and the AAAS-owned *Science 86*, both closed after failing to show a profit.

Because of *Discover*'s unique position in the US magazine market, the magazine was "tremendously important for US science", says Jared Diamond from the University of California, Los Angeles, a contributing editor to *Discover* and a noted science popularizer. He agrees that the commercial performance of *Science Digest* and *Science 86* was disappointing, but maintains that *Discover* has shown it is possible for a science magazine aimed at the general US public to make a profit. Any takers?

Peter Aldhous



advertising revenues. "Discover just wasn't enough to carry everyone else along," says Marc Zabludoff, the magazine's executive editor.

Discover's staff are optimistic that a buyer can be found to relaunch the title. The American Association for the Advancement of Science (AAAS) has been mentioned as a possible buyer by some of the magazine's contributing editors, but

NATURE · VOL 352 · 15 AUGUST 1991

Grand Canyon, clearly

Washington

BY the end of the decade, visitors to the Grand Canyon may get a better view of its spectacular scenery, thanks to a new-found spirit of cooperation between the owners of an Arizona electricity generating plant and their erstwhile environmentalist critics.

The Salt River Project, representing the six co-owners of the Navajo coal-fired station (which include the US Interior Department), last week agreed to submit to the Environmental Protection Agency (EPA) a proposal that would cut the plant's sulphur dioxide emissions by 90 per cent by 1999. The proposal was drafted by the Grand Canyon Trust, an environmentalist group that, together with the US National Park Service, has long argued that the Navajo plant contributes significantly to the haze that can obscure views of the Grand Canyon. The Navajo plant's owners have disputed this claim, and had balked at an earlier EPA proposal to cut the station's sulphur dioxide emissions by only 70 per cent.

The apparent change of heart is due to simple economics. Under the Clean Air Act, EPA was going to force emissions cuts on the plant whether its owners objected or not, and the agency's proposal had been estimated to cost around \$3,000 million, spread over 20 years. The new plan would cost only \$1,800 million over the same period, according to the Salt River Project. Despite greater emissions cuts, it is cheaper because it allows the Navajo plant's operators to measure the cuts on the basis of a yearly average of emissions, rather than the 30-day average proposed by EPA.

Peter Aldhous

BRITISH SCIENCE

WE shall be publishing on 12 September *Nature's* Manifesto for British Science, which will be offered without charge to any one of the political parties contesting the next general election and in the belief that it is an improvement on their own version.

Following publication, on Friday 13 September, there will be a public meeting at the Royal Institution, Albemarle Street, London W1, at which Sir Mark Richmond, chairman of the Science and Engineering Research Council, will be the principal speaker. The chairman will be John Maddox, editor of *Nature*. The meeting will start at 9.45 a.m. and end at 12.45 p.m. The political parties will be invited to send representatives, but there will be ample time for discussion.

Those wishing to attend should apply for tickets to Mary Sheehan, *Nature*, 4 Little Essex Street, London WC2R 3LF.

Imanishi-Kari (continued)

SIR — Wortis *et al.* say (*Nature* 351, 514; 1991) that I could not have known that, at their meeting with Dr Imanishi-Kari on 16 May 1986, they were not shown authentic data on which Table 2 in the disputed *Cell* paper may have been based. Correctly, they say that I was not present. These are the data, also called the June sub-cloning data, 'that are alleged by the draft Office of Scientific Integrity (OSI) report to have been fabricated. So the question is whether the Wortis committee was shown these data, and was misled by them, by some other data or by no data at all.

If the committee had been shown data supporting the claims embodied in Table 2, it is mystifying that it did not say as much in the report requested by the National Institutes of Health (NIH) in 1987, when the validity of Table 2 had already been questioned.

The committee now says that it accepted on 16 May 1986 that the then acknowledged deficiencies of Table 2 were mitigated by the existence of data from unpublished experiments. But did they not have a duty to disclose this circumstance when reporting to the Tufts community and to NIH as well as in their 1988 interviews with the Dingell committee staff?

An explanation is also required (not necessarily from Wortis *et al.*) of this circumstance: by their account, at a meeting arranged at only one day's notice, with no advance warning of the issues to be raised, Dr Imanishi-Kari is said to have produced evidence that convinced the committee that "Dr O'Toole's concerns could be answered"; yet two years later, despite detailed written warning of the questions to be raised, Imanishi-Kari did not produce or even mention to the NIH panel the data that had supposedly allayed the Wortis committee's concerns, but said only after the interview that there were such data.

It is also puzzling that some of the data records produced at the second meeting on 23 May 1986 (at which I was present) have been shown, by indentation analysis carried out by the Secret Service, not to have existed one week earlier. I cannot understand why data that the members of the Wortis committee now say convinced them on 16 May that my concerns were unwarranted should not have been produced for me to see at the meeting a week later, but instead should have been replaced by other data records written during the interval.

I believe that these questions will be answered in due course, and that it will then emerge that the only unpublished data supporting the claims in existence

on 23 May 1986 were those shown to me on that day, and that the more complete answer to my challenge presented to the Dingell committee was fabricated much later.

MARGOT O'TOOLE

Newton,
Massachusetts 02160, USA

Chauvinism

SIR — Publication bias, several forms of which have been described^{1,2}, may distort the scientific literature. On the hypothesis that there is a 'national publication bias' in which journals favour the printing of papers coming from the same country, we have examined all original papers published during 1990 in one British, one Swedish, one US and one German journal. All journals were related to our field of research (physical medicine and rehabilitation) and were listed in *Current Contents*; all were in English, except for one published in Germany, which merely had English abstracts. Each paper was classified according to the address of the first author. The table shows that the nationality of each journal is strongly associated with the nationality of its authors.

PERCENTAGE OF FIRST AUTHORS OF ORIGINAL ARTICLES

Address of journal	National authors	Foreign authors	Impact factor
UK	59.6	40.4	1.6
Sweden	42.4	57.6	0.3
USA	79	21	0.6
Germany	56	44	0.0

This relationship is not uniform. According to the present data, it is strongest in the United States. All other countries follow at considerable and similar distance.

Clearly, these results need to be confirmed on a broader basis using data of many more publications. Yet we suspect that the 'national publication bias' is highly prevalent in the scientific literature. It is of some concern in a world where research should be a free and international matter and could have at least two worrying effects. (1) It might lead to a false representation of scientific opinions according to one nation's preference; if, for example, one form of therapy is neglected in the United States (as calcium channel blockers were several years ago), that would cause a general under-representation of that treatment in the medical literature, which is dominated by US journals with high impact factors. (2) Professional career decisions are now often influenced by impact factors and other measures of scientific output, so that the bias is unfair to

scientists who come from countries where journals rarely have high ratings. If the 'national publication bias' can be confirmed to exist, editors should change their policies and university bodies should think of ways to minimize its harmful effects.

E. ERNST
T. KIENBACHER

Department of Physical
and Medical Rehabilitation,
University of Vienna,
A-1097 Vienna, Austria

1. Easterbrook, P.J., Berlin, J.A., Gopalan, R. & Matthews, D.R., *Lancet* 337, 867-72; 1991.
2. Andersen, B. *Methodological Errors in Medical Research* (Blackwell, Oxford, 1990).

Magnetic machine

SIR — I wish to clarify and correct what Christopher Anderson says (*Nature* 352, 270; 1991) about the susceptometer that this company supplied to the superconductivity Interdisciplinary Research Centre (IRC) at Cambridge.

Far from believing itself bound to buy a British susceptometer from this company rather than the US unit, the IRC bought one of each. Our machine is optimized for weak signals at the lowest magnetic fields. The S100 susceptometer built for the IRC was delivered two years after order but not two years late; it was ordered on a delivery time anticipated to be 14 months.

It is true that the Cryogenic susceptometer was only in the development stage at the time the IRC placed its order, but since then ten systems have been delivered to customers around the world, including the United States and Japan. Once installed correctly, the machines have proved reliable and have been easy to operate and support; customers seem well pleased with the performance.

Not all these machines are used for superconductivity research; the original design produced by both the British and US companies lacked an important feature relevant to research on superconductors. In the normal mode of operation, the sample is moved a relatively long distance, more than 10 cm, and so suffers a change in applied magnetic field during each measurement. This makes it impossible to obtain true magnetic hysteresis loops. To overcome this difficulty, a short scan facility is being provided in which the sample is moved only about 1 cm and remains in a uniform magnetic field. Our company is providing this extra facility free of charge to the IRC at Cambridge, although it was not originally specified. Incidentally, the price was not £100,000 but less than £80,000 for the option chosen by the IRC.

JEREMY GOOD
Cryogenic Consultants Limited,
Metrostore Building, 231 The Vale,
London W3 7QS, UK

Rediscovery of Da Rios equations

Renzo L. Ricca

The remarkable story of the discovery of a set of equations at least three times this century shows once again that independent discoveries can occur and exist for some time.

IN the frenetic activity that seems to characterize modern science, the same discovery may well have more than one parent and, sometimes, more than one life. The following story is a remarkable example of this, for it shows not only that genuine independent discoveries may occur and even today survive independently for some time, but also provides grounds to reflect on common attitudes in science as well as insight into the way in which our scientific culture is formed.

The story begins at the University of Padua in the early years of this century. There, a young Italian, Luigi Sante Da Rios, under the supervision of Tullio Levi-Civita, one of the great mathematicians of this century, had just completed the work for his *Laurea* in mathematics, the Italian degree roughly equivalent to a master of science. The subject of Da Rios' research was fluid dynamics from the point of view of applied mathematics, in particular the study of the free motion of a slender tube-like shaped vortex (a vortex filament) in an unbounded gaseous or liquid fluid medium.

The problems of vortex motion and dynamics of rotational flows had captured imagination since the time of ancient Greece, and throughout the eighteenth and nineteenth centuries many efforts had been made to give these studies rigorous mathematical foundation. At the beginning of this century, then, the vivid influence of works like those done by Lord Kelvin and von Helmholtz upon mathematicians interested in vorticity dynamics was already widely acknowledged. And it was from these studies that Da Rios started to investigate the dynamics of an isolated thin vortex filament as an object embedded in an infinite domain, entirely filled by a homogeneous, incompressible, inviscid fluid. He wanted to study the influence that the localized vorticity has on the local behaviour of the vortex;

to analyse the kinematics and the dynamics of the vortex filament so as to understand its motion; and, eventually, to apply these results to examine its global behaviour.

By assuming the thinness of the vortex core and the local effects due to vorticity as relevant hypotheses to his mathematical model, Da Rios derived the related asymptotic potential theory to find a simple approximate expression for the self-induced velocity of the vortex. This law says that the vortex filament moves

in the intrinsic parameters. One solution describes the configuration of a plane vortex that moves with a 'catenary'-like shape, a shape taken under the gravitational force by a flexible, inextensible thread of uniform weight suspended at two points. The other describes the configuration of a space vortex moving with a particular three-dimensional shape like a travelling wave.

All these results were collected by Da Rios, then 25 years old, in a paper¹ entitled "*Sul moto d'un liquido indefinito*

con un filetto vorticoso di forma qualunque" (On the motion of an unbounded fluid with a vortex filament of any shape), printed in 1906 in the July–August issue of *Rendiconti del Circolo Matematico di Palermo*, an international mathematical journal in which, at that time, mathematicians like Cantor, Hilbert, Poincaré, Peano and many others used to publish. Unfortunately, the paper didn't bring much recognition to Da Rios. Only Levi-Civita, who already knew the results, gave the author the merit due to him. In a paper published in 1908², Levi-Civita, re-examining the whole

SUL MOTO D'UN LIQUIDO INDEFINITO CON UN FILETTO VORTICOSO

che, per le (18) e (21), diventano:

$$-\frac{d\tau}{dt} - \left(\frac{c'}{c} - \tau^2\right)' = cc',$$

$$c'' = c'' - c\tau^2 + c\tau^2,$$

$$\frac{dc}{dt} = c\tau' + 2c'\tau.$$

Abbiamo quindi finalmente le equazioni cercate:

$$(22) \quad \begin{cases} \frac{dc}{dt} = c\tau' + 2c'\tau, \\ \frac{d\tau}{dt} = -cc' + \left(\tau^2 - \frac{c''}{c}\right)'. \end{cases}$$

Il teorema di esistenza, applicato a questo sistema di equazioni a

mette di asserire con tutto rigore che le funzioni $c(s, t)$, $\tau(s, t)$

(regolarità) univocamente definite dai valori inizi

The intrinsic equations (22) as they were presented by Da Rios in his first paper published in 1906; c and τ stand for curvature and torsion of the vortex filament, respectively.

with a velocity proportional to its local curvature: if we take two vortex rings (a smoke ring is a good example) of equal internal structure and equal core size, but of different diameter, then the smaller ring will move faster. The set of basic assumptions that leads to this law of motion is nowadays commonly referred to as the localized induction approximation (LIA). Using LIA, Da Rios went further and studied the global behaviour of the space vortex filament. He found a remarkable system of two coupled equations, hereafter referred to as the Da Rios equations, which governs the motion of the vortex in terms of time evolution of its geometric intrinsic parameters, curvature and torsion. He completed his research by demonstrating the existence of two physical solutions to these equations describing rigid motion

work on the asymptotic theory from another point of view, states: "For a long time in hydrodynamics, in the study of rectilinear or circular vortex filaments, we had to deal with special asymptotic expressions. But Da Rios has the merit of being the first to establish a general approach of this kind" (the general asymptotic theory and the LIA velocity). At that time Levi-Civita was interested in the gravitational attraction produced by a slender filament of mass for possible applications to the saturnian ring, and he worked for several years on improvements of the asymptotic potential theory and on its applications³⁻⁵.

In the meantime, Da Rios became assistant lecturer at the University of Padua; there, very probably, the daily discussions he would have had with Levi-Civita, who is known to have been

enthusiastic about his work, convinced him to publish his results once again. He prepared two versions^{6,7}: one, the shorter, was presented by Levi-Civita himself for publication, and published in the *Rendiconti della Reale Accademia dei Lincei*; the other, written as a review article, appeared in 1910, in the spring issue of *Rendiconti del Circolo Matematico di Palermo*. In both papers, Da Rios explained once again in detail the asymptotic potential theory he had developed, the LIA approach and the intrinsic equations of motion. Still, for some unknown reason, both works remained obscure, and Da Rios turned his attention to other aspects of vortex dynamics. Levi-Civita, on the other hand, was apparently so fascinated by the Da Rios results that in 1932 he collected them in a survey⁸, improving and extending the Da Rios analysis in many directions. He studied in detail the stability of the circular vortex, the case of a helicoidal vortex filaments and a particular plane configuration, characterized by a knot-like shape, for a vortex moving without change of form; recalling what Da Rios had briefly sketched, he analysed in the intrinsic parameters the theory of rigid vortex motion.

Many years passed, and this last work of Levi-Civita also fell into obscurity. To understand this lack of interest it is useful to remember that at the beginning of the century, and for several decades thereafter, the problems considered fundamental in fluid dynamics were the structure and the role of the boundary layer produced by a fluid flowing along a solid surface, the formation of wakes behind obstacles, and the study of the fully developed turbulence. It is not surprising, then, that more than 50 years passed before the localized induction concept was rediscovered, when Hama and coauthors⁹⁻¹¹, unaware of Da Rios's work, proposed the LIA approach in several papers published in the well-known US journal *Physics of Fluids*. It is common to refer to Hama as having formulated LIA; nonetheless in 1962 he acknowledged Arms for having conceived the LIA concept¹⁰.

An immediate consequence was the independent rediscovery of the Da Rios equations in fluid dynamics, in 1965, by Betchov¹², who also found the rigid plane vortex solution originally derived by Da Rios in 1906 and later studied by Levi-Civita. A few years later, by coupling curvature and torsion of the filament into one complex variable, Hasimoto¹³ combined the two intrinsic equations (then termed the Betchov equations) into the nonlinear Schrödinger equation, one of the 'soliton' equations now much studied in the mathematics of nonlinear systems, such as optics, acoustics and signal-transmission theory.

Soliton equations give travelling wave-like solutions: on a space curve, for example, soliton solutions are three-dimensional pulse-like solitary waves travelling on the curve, asymptotically preserving their shape and velocity upon collision with other solitary waves. Under LIA, one of these travelling waves is given by the three-dimensional physical solution found by Da Rios.

Sometimes theories about different aspects of physics may not differ much in their mathematical structure; exact or formal analogies can occur. Analogies are useful in science, because if we deal with the same mathematical model applied to different physical contexts, and are looking in the same direction, we are likely to find the same things, even if we call them by different names. This is true in the case of Lakshmanan and his colleagues¹⁴, who studied, in the context of ferromagnetic media, the dynamics of a one-dimensional system of classical spins — that is a space curve on which the spin vectors, in the continuum limit, lie. Under certain assumptions it is possible to show that the motion of such an object is governed according to LIA. By taking this into consideration, Lakshmanan and colleagues derived, for the third time, the Da Rios equations and found the solitary wave solutions. What is surprising, though, is that when Lakshmanan presented these equations again in a brief letter in 1977¹⁵, he mentioned Betchov and Hasimoto but did not reference the equations that Betchov had earlier (re)discovered.

A whole family of travelling wave solutions, initially investigated by Levi-Civita in his paper of 1932, were found successfully by Kida¹⁶ in 1981, and the soliton behaviour of the related nonlinear Schrödinger equation began to be well known. Of the soliton behaviour of his intrinsic equations Da Rios was of course unaware; nonetheless when he studied the three-dimensional pulse-like travelling-wave solution he entitled that part of his paper "Research of hunched vortices travelling rigidly".

It was only at the beginning of the 1980s that Germano became aware of Da Rios' work and wrote a paper¹⁷ in which he reaffirmed the discoveries of Da Rios and extended the original set of

intrinsic equations by considering the full general kinematics of a space curve, independently of the particular physical context. This paper, presented in 1983 at an Italian congress, was translated into English, but unfortunately this version was never published. Da Rios became widely known only when Moffatt¹⁸, informed by Germano about his work, acknowledged him in his talk at a symposium in 1983 held in Kyoto, this time in front of an international audience and, of course, in English!

But surprises are still possible. Recently, Hama¹⁹ gave his account of the genesis of LIA: talking about his 'theorem', he said that according to Hasimoto, LIA was formulated in Tokyo in 1937²⁰. And Keener²¹ last year derived again the set of general intrinsic equations, as Germano had done in 1983, acknowledging Betchov and Germano only in a note added in proof.

With the growth of interest in exotic string objects — classical, cosmic, super or whatever — and with the application of the mathematics of nonlinear systems in many fields of physics, chemical physics and biological physics, attention given to the Da Rios-Betchov equations has increased greatly. Although connections between the intrinsic kinematics of curves and the general properties of one-dimensional soliton equations have been brought into evidence²², many questions remain open. Recently, further generalizations of the intrinsic equations to higher-dimensional cases have been proposed²³, which may be interesting not only because the underlying mathematical structure of the equations could provide further evidence of links to more general sets of non-linear equations, but also because many physical problems might find in this approach their suitable formulation. In pure mathematics, too, these equations can be of interest: various problems in the geometry of string-generated surfaces²⁴, await solution. My own feeling is that the story of the Da Rios equations is not yet finished. □

Renzo L. Ricca is in the Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge CB3 9EW, UK.

1. Da Rios, L. S. *Rend. Circ. Mat. Palermo* **22**, 117–135 (1906).
2. Levi-Civita, T. *Rend. R. Acc. Lincei* **17**, 3–15 (1908).
3. Levi-Civita, T. *Rend. R. Acc. Lincei* **17**, 413–426, 535–551 (1908).
4. Levi-Civita, T. *Rend. R. Acc. Lincei* **18**, 41–50 (1909).
5. Levi-Civita, T. *Rend. Circ. Mat. Palermo* **33**, 354–374 (1912).
6. Da Rios, L. S. *Rend. R. Acc. Lincei* **18**, 75–79 (1909).
7. Da Rios, L. S. *Rend. Circ. Mat. Palermo* **29**, 354–368 (1910).
8. Levi-Civita, T. *Attrazione Newtoniana dei Tubi Sottili e Vortici Filiformi* (Annali R. Scuola Norm. Sup. Pisa, Zanichelli, Bologna, 1932).
9. Hama, F. R. & Nutant J. *Phys. Fluids* **4**, 28–32 (1961).
10. Hama, F. R. *Phys. Fluids* **5**, 1156–1162 (1962).
11. Arms, R. J. & Hama, F. R. *Phys. Fluids* **8**, 553–559 (1965).

12. Betchov, R. J. *Fluid Mech.* **22**, 471–479 (1965).
13. Hasimoto, H. J. *Fluid Mech.* **51**, 477–485 (1972).
14. Lakshmanan, M., Ruijgrok, T. W. & Thompson, C. J. *Physica* **A84**, 577–590 (1976).
15. Lakshmanan, M. *Phys. Lett.* **A61**, 53–54 (1977).
16. Kida, S. J. *Fluid Mech.* **112**, 397–409 (1981).
17. Germano, M. *Proc. VII natn. Cong. Assoc. Ital. Aeronaut. Astronaut.* **1**, 163–170 (1983).
18. Moffatt, H. K. in *Turbulence and Chaotic Phenomena in Fluids* 223–230 (Elsevier, Amsterdam, 1984).
19. Hama, F. R. *Vortex Motion — Fluid Dynamics Research* **3**, 149–150 (North-Holland, Amsterdam, 1988).
20. Murakami, Y., Takahashi, H., Ukita, Y. & Fujihara, S. *Oyo Butsuri* **6**, 151–153 (1937).
21. Keener, J. P. J. *Fluid Mech.* **211**, 629–651 (1990).
22. Lamb Jr., G. L. J. *Math. Phys.* **18**, 1654–1661 (1977).
23. Ricca, R. L. *Phys. Rev. A* **43**, 4281–4288 (1991).
24. Sym, A. *Let. Nuovo Cimento* **41**, 353–360 (1984).

NIH need clear definition of fraud

Behind the pyrotechnics surrounding the way the National Institutes of Health investigate fraud are substantive issues of fairness.

THE 1 August hearing in the US Congress at which Representative John Dingell locked horns with Bernadine Healy, the new director of the National Institutes of Health (NIH), became a battle of titans as Healy boldly stood her ground in the face of Dingell's accusations that she was dismantling the NIH's fraud office for personal and vindictive reasons (see *Nature* 352, 461; 8 August 1991).

At issue was Dingell's assertion that Healy had no business getting into the middle of the Office of Scientific Integrity's interminable investigation of two cases involving, separately, virologist Robert C. Gallo and Nobel laureate David Baltimore and Thereza Imanishi-Kari. Dingell was enraged that Healy had taken steps to 'rein in' OSI's former acting director, Suzanne Hadley, who she said was editorializing inappropriately in the OSI's official reports and who was suspected of becoming too close to the whistleblower in the Baltimore case. Dingell said that because Hadley had reviewed and found wanting Healy's own handling of a case at the Cleveland Clinic, Healy's previous employer, Healy could be accused of a conflict of interest — a charge Healy called "preposterous".

Behind the colourful pyrotechnics of the hearing lie several substantive issues crying out for resolution in a serious and sober environment. The first is the government's very definition of misconduct, which is vague and open-ended to the point that OSI has been at liberty to label as wrongdoing any behaviour that it decides does not fit the 'norm'.

The second is the question of whether Dingell's continuous pursuit of alleged fraud has become so dominant and distracting that potentially life-saving science is going undone. Martin Delaney, head of Project Inform, one of the largest US AIDS activist groups, has taken that tack and threatens to launch a vigorous effort to stop what he sees as Dingell's harassment of scientists.

Finally, a status report by the government's Office of Scientific Integrity Review, which has authority over NIH's fraud office, suggests that the incidence of serious wrongdoing is, indeed, rather minor. For instance, of 21 cases reviewed by OSIR as of December 1990, findings of misconduct are broken down as follows: data fabrication or falsification, six; plagiarism, five; "other deviant practices", such as multiple manuscript submission, seven. With respect to the five cases of plagiarism, the report says each involved the use of review material that was acknowledged in a bibliography but not sufficiently credited. "There were no cases where research results or research data were found to be plagiarized."

It is this kind of information that makes one wonder whether the NIH are spending too much time investigating trivia, and it is the ill-defined matter of "other deviant practices" that has Healy demanding a new look at the institutes' approach to policing science. It is here that Healy's quarrel with Hadley lies.

In 1986, official guidelines defined misconduct as fabrication, falsification and plagiarism. Three years later the definition was amended to include "other practices that seriously deviate from those that are commonly accepted within the scientific community for proposing, conducting, or reporting research". Healy wisely would like to see the latter go, leaving the government's fraud investigators to concentrate on real fraud, committed with intent. "Intent and purposefulness", in Healy's view, are necessary components of fraudulent behaviour — the element that distinguishes fraud from sloppiness or error. Hadley, on the other hand, has explicitly told scientific groups that she finds intent irrelevant in identifying misconduct.

Healy is also taking issue with OSI's definition of due process which expressly denies the accused the right to see at first hand all of the evidence against him or her. Hadley's organizing principle for OSI has been that if someone sees all of the evidence, he or she is in a better position to explain it away. Better to keep the allegedly damaging data secret, available to the accused only in summary form.

When it comes to procedure, due process is not the only issue. Healy reports her "astonishment" at witnessing a debate amongst high-level NIH staff about just which set of procedures the fraud office is supposed to be following. She finds the OSI's definition of conflict of interest at odds with that used by the rest of NIH. Healy is distressed that Hadley apparently failed to follow existing guidelines for logging phone calls and blocked the early release of the Gallo report to an advisory committee that was not duly constituted. Healy recalls early

advice from Dingell: to do a good job, follow the rules. Healy wants OSI to have good, clear rules.

Then there is the matter of confidentiality. Healy is furious that the OSI's draft report in the Imanishi-Kari case was widely leaked to the press where its preliminary findings were often reported as final. Add to that what is suspected to be OSI's recent leak of a draft report in the Gallo case (see page 555) and Healy has every reason to think that the office has "run amok".

A fundamental point needs to be resolved in this debate. Is it the government's job to ferret out and punish scientists who commit fraud in the course of conducting federally funded research, or, should the government's official fraud office extend its reach to matters that are properly defined as error and carelessness? A logical response is that the government should stick to the former, leaving judgements about the adequacy of footnotes, perfection of data presentation in published tables or handling of students to editors, tenure committees and the bodies that award prizes to people whose behaviour is so exemplary that it sets them above the average.

Concern that the government is getting into the business of bringing its full weight to bear on ordinary human imperfections, whether through NIH or the office of Representative Dingell, is beginning to have a psychologically crippling effect on the research community. It is that angst that AIDS activist Delaney hopes to tap. In a letter to Dingell, Delaney accuses the congressman of acting as judge and jury at the Healy hearing before even hearing her testimony and finds "a pattern of bias in your office stretching back for more than three years".

In a new twist to the story, Delaney is calling for an independent investigator of the sort brought into the Watergate and Iran-Contra scandals to examine not only the Office of Scientific Integrity but also Dingell's office itself.

"Bureaucratic zeal", Delaney charges, "is itself a major enemy in the fight against AIDS." A petition signed by AIDS activists, community figures and, he hopes, scientists will mark Delaney's first move in this new battle to get research instead of fraud at the top of the country's scientific agenda.

Barbara J. Culliton

A superplume in the mantle

K. G. Cox

THE Earth's mantle, R. L. Larson suggests in *Geology*¹, may have undergone a vast hiccough in circulation 120 million years ago, the last effects of which may still be apparent in the Tahiti 'super-swirl'. Perhaps related to events in the Earth's core, the hiccough, Larson argues, caused an outburst of production of oceanic crust, and it may even have been responsible for global climate warming at the time.

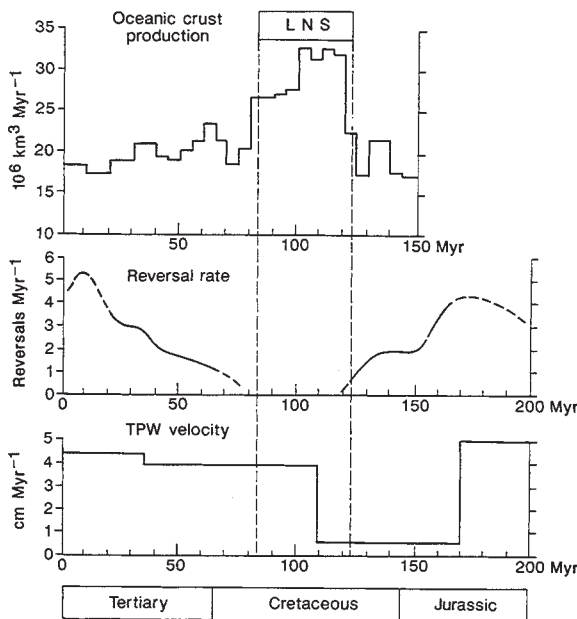
The idea that rising hot jets of mantle material (plumes) constitute an important, even possibly the most important, component of the convection of the Earth's mantle, has been growing in popularity among earth scientists since its first appearance nearly 30 years ago. The original concept² sought to explain chains of volcanic islands which appeared to migrate across oceanic plates. These were interpreted as representing the volcanic expression of the movement of plates across relatively fixed hotspots or plumes — in other words, the mantle, as well as undergoing convection associated with plate movements, contains a long-lived, and probably deep-seated, stable convective pattern of plumes, over which plates independently move. The plumes were, by implication, pencil-like in form — narrow jets giving rise to narrow chains of islands such as the Hawaiian islands and the Emperor sea-mount chain.

Then evidence of larger plumes, or at least with much more extensive effects at the surface, began to emerge, so that by 1989 some plumes were postulated as having active heads 2,000 km across³, explaining large continental flood basalt provinces and oceanic plateaux. R. L. Larson's new paper¹ carries this trend one step further, in identifying a so-called superplume which affected areas mainly in the Pacific during the mid-Cretaceous (120–80 million years ago) and had effects over lateral distances of perhaps 6,000 km, a significant fraction of the circumference of the Earth.

Larson's hypothesis is based on a careful estimate of variations in the rate of production of new oceanic crust with time, over the past 150 million years. He finds a 50–75 per cent increase in the production rate during the mid-Cretaceous (top part of figure), which he attributes to the activity of a superplume, originating at the core–mantle boundary about 125 million years ago,

and still perhaps represented, though in a near-exhausted state, by the present-day 'superswell' under Tahiti.

If superplumes do exist, however, they are not likely to be manifestations of just mantle convection. Convection on this scale, if originating in the lower mantle, may be related to events in the core, such as reversals of the Earth's magnetic field. As evidence of this, Larson demonstrates that the period of maximum oceanic crust production coincides very



Variation with time of (top) estimated world oceanic crust production rate (excluding Tethys)¹, (middle) magnetic reversal rate⁵, and (bottom) velocity of true polar wander⁵. Vertical broken lines enclose the long normal superchron (LNS).

closely with the long period of normal polarity (the 'long normal superchron' or LNS) during the Cretaceous, although he does not speculate in detail about the physics of the core–mantle interaction. But the implications of superplumes do not stop here. If unusually large convective and melting events take place, the resultant surface volcanism, involving massive additions of CO₂ to the atmosphere, may have dramatic effects on climate, and on the biosphere in general⁴.

The general hypothesis that mantle convection — the main mechanism of heat loss from the Earth's interior — may control major events both at the surface and in the core has always been exciting and plausible. But what is the nature of the new evidence, how compelling is it, and indeed, in terms of general scientific method, how should such speculations be conducted?

Larson's calculations of rates of pro-

duction of oceanic crust are based on a number of clearly stated assumptions. Little is known, for example about the Cretaceous crust of the eastern Pacific, most of which no longer exists, having been subducted under the Americas. Symmetrical spreading has to be assumed for an estimate of the missing volume. It also has to be assumed that most of the large oceanic plateaux of the western Pacific (which have much thicker crust) had eastern counterparts that are no longer available. Similarly, nothing can be said about production of oceanic crust, if any, in the vanished ocean of Tethys. All the same, the calculated increase in crust production

during the mid-Cretaceous is so large that the conclusions deserve serious attention.

More problematic is the question of whether crust production at mid-ocean ridges (as opposed to oceanic plateaux) is a safe measure of the vigour of deep-seated mantle convection. The amount of sea-floor spreading at a particular period clearly may be influenced by plumes, but it is also a function of the amount of concurrent subduction. Larson argues that the latter did not vary significantly during the period in question, but the evidence is difficult to assess.

This raises the question, what is the best thing to measure if we wish to monitor deep-seated convective activity in the mantle? Perhaps in the future, using Larson's type of approach, detailed knowledge of basalt geochemistry may permit the argument to be restricted to the volume of volcanism suspected to be purely plume-related (and that would include continental flood basalts), while leaving out production at ridges. Courtillot and Besse⁵ have taken a quite different approach, however, using the velocity of true polar wander (TPW — the migration of the hotspot reference frame in the mantle relative to the spin axis of the Earth) as their indicator of important changes in mantle convection. In their analysis (middle and bottom parts of the figure), they found a period of very slow TPW from about 170 to 110 million years ago, coinciding with a significant burst of continental break-up and largely preceding the mid-Cretaceous LNS.

Their important correlation was thus between very slow TPW, continental fragmentation (which was endemic during the same period), and the fairly consistent diminution in the geomagnetic reversal rate during the late-Jurassic and early-Cretaceous. Then, shortly after the mid-Cretaceous LNS had started, TPW velocity increased suddenly to more-or-

less its present high rate. Not surprisingly, their model of core-mantle interaction shows some significant differences from Larson's, which directly relates heat loss from the lower mantle to simultaneous stabilization of the core. From their evidence, slow TPW — that is, reduced convective activity — corresponds to a decreasing reversal rate. Conversely, fast TPW is accompanied by an increasing reversal rate, so, as in Larson's model, increased mantle activity is somehow related to core stabilization. But as the figure shows, some substantial time-lags have to be built into Courtillot and Besse's model, with events in the mantle significantly preceding changes in the core.

There is no doubt that the issues under discussion are of supreme importance to the earth sciences. If the coupling of core, mantle and surface events (including climatic) can be firmly established, a substantial unifying framework will be available for the study of most major geological processes, though events of extraterrestrial origin may yet muddy the waters. The comparison of Larson's and Courtillot and Besse's papers, however, illustrates some of the obstacles to further progress.

First, at present it is permissible to postulate many different things about the physical nature of happenings at the core-mantle boundary, within the mantle itself, and within the core. The generally acceptable theoretical models, against which the observational facts ultimately have to be tested (because many of the phenomena are outside the range of direct observation), do not yet exist in sufficient detail. Superplumes demand supercomputers.

Second (within the observational field) is the problem of correlation of events in time, which is so often (and not always wisely) the basis of geological hypothesis. Larson, for example, demonstrates a superb one-to-one correlation, but unfortunately there is only one period of increased production of oceanic crust to correlate with one period of geomagnetic stability. What is the level of significance? Ultimately, if time correlation is the key to the argument, we need to define as many separate volcanic events as possible, or whatever other events are used to measure mantle convective activity. □

K. G. Cox is in the Department of Earth Sciences, University of Oxford, Parks Road, Oxford OX1 3PR, UK.

Disease and evolution

J. C. Howard

"Now every species of mammal and bird so far investigated has shown a quite surprising biochemical diversity revealed by serological tests. The antigens concerned seem to be proteins to which polysaccharide groups are attached. We do not know their functions in the organism, though some of them seem to be part of the structure of the cell membrane. I wish to suggest that they may play a part in disease resistance, a particular race of bacteria or virus being adapted to individuals of a certain range of biochemical constitutions, while those of other constitutions are relatively resistant" (J. B. S. Haldane, 1949; ref. 1).

THIS passage is typical Haldane — clear, original, synthetic, precocious. But was he right? Haldane was aiming to explain not the existence of membrane-bound glycoproteins as such, but rather their polymorphic diversity. By virtue of their overwhelming advantage in rate of evolution, microbes seemed always to have the upper hand against host attempts to evolve a resistance mechanism common to all members of the species. Haldane conceived of a form of equilibrium in which host resistance factors vary from individual to individual. Infectious pathogens would then tend to learn to deal with the commonest type, leaving the way clear for rare types to persist, and eventually flourish for a time.

The membrane glycoproteins coded by the major histocompatibility complex (MHC) were some of the main contributors to Haldane's "quite surprising biochemical diversity". We now know that the MHC indeed seems to function as a resistance system, apparently exactly as Haldane conceived it. Polymorphic glycoproteins encoded by the MHC play a central role in the immune system; peptide fragments of infectious origin can be captured in the peptide-binding grooves of both class I and class II MHC molecules and presented by them to the T-cell immune system. Furthermore, the astonishing polymorphism of MHC molecules largely centres upon those residues which interact with peptides², and the large excess of coding substitutions at exactly these positions³ suggests, exactly as Haldane's theory predicts, that there is strong selective pressure in favour of variation.

All this is good physiology and good genetics, but does it constitute proof of Haldane's theory? Surely, what is needed is explicit evidence that MHC allele frequencies are driven by pathogens? On pages 595 and 619, *Nature* now publishes two outstanding

studies on the selective pressures operating on MHC polymorphism in the wild^{4,5}, and they reach what at first sight seem to be opposed conclusions, one wholly for Haldane, the other wholly against. I shall try to reconcile these two studies, and will argue that they should in fact both be seen as strengthening rather than weakening the generalization that infectious disease is the principal motor for polymorphism.

First, for Haldane, Hill and colleagues⁴ face the problem directly. Find a disease associated with a high mortality in the wild, a disease having a high prevalence and in a species in which it is possible to identify most histocompatibility alleles with precision. In addition, it must be possible to assess disease-related mortality at all stages in the life cycle. The species is the challenge of course. Only the human qualifies. The disease is falciparum malaria, hyperendemic in West Africa, a potent killer of the young and already established as the exemplar for natural selection through its effect in elevating the frequency of HbS, the sickle cell haemoglobin allele⁶ (still, after more than 30 years, the only balanced polymorphism in man for which we have an explanation).

Hill *et al.* compared human leukocyte antigen types in children desperately ill with severe malarial anaemia or cerebral malaria against a number of control groups. This was a brave and correct analysis, for the story is about natural selection, and natural selection in this context means life-threatening illness. Out of 45 class I alleles (assayed first by serology, then confirmed in a different sample by the polymerase chain reaction), one alone, HLA-Bw53, was significantly reduced in frequency in the severely ill children. The hypothesis that HLA-Bw53 is indeed a protective allele against the severest forms of malaria is confirmed by its geographical distribution; vanishingly rare elsewhere in the world, HLA-Bw53 reaches a frequency as high as 25% in malarial regions of Africa. Out of 13 class II haplotypes assayed, again one, DRB1*1302-DQB1*0501, was significantly reduced in frequency in one category of the severely ill children, namely those with severe malarial anaemia. The class I and class II protective mechanisms are presumably different, and suggest new approaches to vaccine development.

Although the degree of protection conferred on an individual by the protective HLA alleles was less than that provided by HbS, the higher frequency

1. Larson, R. L. *Geology* **19**, 547–550 (1991).
2. Wilson, J. T. *Can. J. Phys.* **41**, 863–870 (1963).
3. White, R. S. & McKenzie, D. P. *J. geophys. Res.* **94**, 7685–7730 (1989).
4. Caldeira, K. & Rampino, M. R. *Geophys. Res. Lett.* **18**, 987–990 (1991).
5. Courtillot V. & Besse J. *Science* **237**, 1140–1147 (1987).

of the HLA alleles means that, overall, their protective effect must be greater, representing a total equivalent to about 15% of all severe malaria cases in the Gambia, against 12% for HbS. The population of West Africa has been dense enough for efficient malarial transmission for no more than 10,000 years; presumably HLA-Bw53 has reached its 25% frequency from close to zero in that time by simple directional selection. Its now-known selective advantage in a malarial environment is easily large enough to achieve this incidence.

The premise of Haldane's theory is that parasites always out-evolve their hosts. Will *Plasmodium falciparum* therefore soon take avoiding action? If not, should we expect HLA-Bw53 to carry on increasing in frequency at the expense of other HLA alleles? Presumably (but it now seems a fair presumption), other pathogens will take advantage of Bw53 as its frequency rises, though whether, in a modern medical context, we shall ever see the consequence is open to doubt. In any event, the results of Hill *et al.* clearly demonstrate the essential validity of Haldane's claim, and reinforce other evidence in its favour both in human infectious disease⁷, and also, in one well-documented case, in chickens^{8,9}.

The argument for infectious disease as the driving force for MHC polymorphism now seems overwhelming. So when Potts and colleagues⁵ established populations of half-wild mice carrying known histocompatibility types at known frequencies in an enclosed but naturalistic environment, they can hardly have anticipated what would happen. Essentially right away, genotype frequencies departed radically from expectation, with far too many MHC heterozygotes being found in the progeny. Were MHC homozygotes peculiarly susceptible to infectious disease, and dying young? Potts *et al.* showed that this was definitely not so, and that the whole excess of heterozygotes arose from mating choices made by the females. Not only did females initially tend to establish territorial relationships with males differing at the MHC, but they also devoted more than half of their reproductive activity to contracting extraterritorial liaisons with further MHC-incompatible males. So strong was the disassortative mating that Potts *et al.* were able to demonstrate that this force, by itself, could be responsible for maintaining the known polymorphism of MHC alleles in the wild mouse population.

The ability of mice and rats to use MHC alleles as individuality cues has

been known for some time^{10,11}. Different MHC alleles are associated with distinct urinary odours which are a basis for mating preference, though the pattern of preference has not been formally established. It is striking that the effect of the MHC should so predominate, in Potts *et al.*'s study, that it could not be overridden by segregation at the many hundreds of other loci by which the mice must have differed from each other.

What does this result mean? It can be reconciled with Haldane's theory by



A Spanish cock in full plumage, with a fine fat comb and wattles (from Darwin's *Variation of Animals and Plants under Domestication*). A lot of what is on display between animals indicates health status — 'feeling off colour', 'bright-eyed and bushy-tailed', 'in the pink of condition', these phrases all tell us how quickly we perceive disease and health. Infectious disease may drive sexual selection too¹⁸.

Potts *et al.*'s own suggestion that female disassortative mating preference, by ensuring MHC heterozygosity, confers enhanced fitness on the progeny through increased disease resistance. Their alternative proposal is that the MHC is serving as a polymorphic marker in a genetic incompatibility system for avoiding the generalized deleterious consequences of genome-wide inbreeding. One should note, however, that the first

hypothesis is in fact a special case of the second, where the relevant locus for inbreeding avoidance is the MHC, and the deleterious consequences are decreased resistance to infectious disease. Furthermore, although a single-locus marker for genome-wide diversity can work, it is most effective at loci linked to, or identical to, the incompatibility locus itself. The simple teleology of the matter therefore suggests that the MHC is itself the genomic locus whose heterozygosity is of principal evolutionary concern.

So we return to infectious disease, but now at one remove. Not only does infectious disease apparently directly drive polymorphism in the MHC by differential mortality, but we must now conclude that it also regulates mating preference in order to maximize heterozygosity at these all-important loci for disease resistance. Heterozygotes may well flourish at the expense of homozygotes, as a direct result of the ravages of disease; however, by disassortative mating, wild mice seem to be achieving the same goal at a lower cost. Not only that, but rodents may even have hijacked the infectious micro-organisms themselves in order to provide the volatile urinary odorants that serve to identify MHC alleles¹². There is evidence that the urine of germ-free rats of different genotype cannot be distinguished¹³, but contrary results have been obtained in mice¹⁴ and the relevance of these observations to MHC-related mating preference is yet to be established. If the pressure of infectious disease can influence olfactory preference in mating, perhaps it can affect other cues as well. It no longer seems implausible to suppose that well-developed secondary sexual characters (see figure) may primarily indicate freedom from infectious disease.

Haldane saw, of course, that the selection pressures of infectious disease would be strong enough to bring other adaptation in tow. He proposed, for example, that polymorphic resistance loci would tend to have high mutation rates. Is it an accident that the highest

- Haldane, J. B. S. *La Ricerca Scientifica* **19** Suppl. 19, 68–75 (1949).
- Bjorkman, P. J. *et al. Nature* **329**, 512–518 (1987).
- Hughes, A. L. & Nei, M. *Nature* **335**, 167–170 (1988).
- Hill, A. V. S. *et al. Nature* **352**, 595–600 (1991).
- Potts, W. K., Manning, C. J. & Wakeland, E. K. *Nature* **352**, 619–621 (1991).
- Allison, A. C. *Br. med. J.* **1**, 290–294 (1954).
- de Vries, R. R., Khan, P. M., Bernini, L. F., van Loghem, E. & van Rood, J. J. *Immunogenet.* **6**, 271–287 (1979).
- Pazderka, F., Longenecker, B. M., Law, G. R. J., Stone, H. A. & Ruth, R. F. *Immunogenetics* **2**, 93–100 (1975).
- Briles, W. E., Stone, H. A. & Cole, R. K. *Science* **195**, 193–195 (1977).

- Yamazaki, K. *et al. J. exp. Med.* **144**, 1324–1335 (1976).
- Singh, P. B., Brown, R. E. & Roser, B. J. *Nature* **327**, 161–164 (1987).
- Howard, J. C. *Nature* **266**, 406–408 (1977).
- Singh, P. B. *et al. J. chem. Ecol.* **16**, 1667–1682 (1990).
- Yamazaki, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **87**, 8413–8416 (1990).
- Klein, J. *Adv. Immun.* **26**, 55 (1978).
- Hamilton, W. D. J. *Genet.* **69**, 17–32 (1990).
- Hamilton, W. D., Axelrod, R. & Tanese, R. *Proc. natn. Acad. Sci. U.S.A.* **87**, 3566–3573 (1990).
- Pomiankowski, A. N. *Oxf. Surv. evol. Biol.* **5**, 136–184 (1988).

recorded mammalian mutation rate is in an allele of a class I gene from the mouse MHC (ref. 15), probably achieved by a specialized mutational mechanism? He did not, however, refer to the ultimate extension of his position, namely, that although infectious disease may be the principal cause of polymorphism, it would be virtually ineffective in the absence of segregation and, for multiple loci, independent assortment. In other words, the sexual POPULATION GENETICS

mechanism itself is an essential ingredient of the resistance system. Which came first? Most of us have been brought up with the idea that sex causes disease. A longer term view, and one for which there is increasing support, is that the boot belongs on the other foot^{16,17}. □

J. C. Howard is in the Department of Immunology, AFRC Institute of Animal Physiology and Genetics, Babraham, Cambridge CB2 4AT. UK.

A way to world knowledge

Jared M. Diamond

THINK of what we could deduce if we knew the DNA base-pair sequence of every human now alive. We could reconstruct much of the history of modern human populations in terms of prehistoric migrations and founder effects. We could study the social structure of human populations, the frequencies and types of human mutations, and natural selection in humans along environmental gradients. We could understand our differing genetic susceptibility to many diseases. Although that goal of completely sequencing everyone is unfortunately not feasible, a flurry of papers shows that a much lesser enterprise could yield a great deal of information¹⁻³.

The immediate impetus comes, of course, from the Human Genome Project, which seeks to sequence completely what amounts to one Caucasian's genome. That endeavour does not in itself address the questions outlined above, but Cavalli-Sforza *et al.*¹ argue that with a mere one per cent of extra effort and expense it could. Two practical issues arise: how to allocate resources among populations (for example do we learn more by sequencing the genomes of 500 individuals from each of 20 populations, or those of 20 from each of 500?); and how large a fraction, and which specific fractions, of each individual's genome to sequence. If we had already sequenced everybody, we could then with hindsight design an efficient sampling strategy. The dilemma is that we don't know the results, yet we have to guess at them to design the strategy.

An essential consideration that Cavalli-Sforza *et al.* point out stems from the main pattern of human history for the past 10,000 years. By virtue of acquiring plants and animals suitable for domestication, certain groups that formerly constituted only tiny fractions of the world's human population have overrun much of the globe. Prime examples are the expansions of Anatolian farmers and proto-Indo-European speakers over Eurasia, of Austronesians

over Indonesia and the Australian realm, of Bantus over sub-Saharan Africa, and of modern Europeans over Australia and the Americas. Conversely, other populations that once accounted for much of human diversity and population numbers are now close to vanishing, overrun by the expansions of others. Cavalli-Sforza *et al.* answer the first of the two practical issues by proposing a crash programme for collecting samples (permanent cell lines, DNA or both) from these vanishing populations. With that material secured, the matter of how to study it could then be addressed at leisure.

Any geneticist or anthropologist could quickly come up with a wish-list of declining human populations to sample. High on anyone's list would come the !Kung of Namibia and Botswana, remnants of the population that occupied most of southern Africa until a few millennia ago, and located at the deepest branches in the world tree of mitochondrial DNA (mtDNA)⁴; scattered tribal peoples of southeast Asia and Indonesia, probably remnants of that area's former population related to native Australians and Melanesians and swamped by the Austronesian expansion; and Japan's Ainu, remnants of who-knows-what. Other relict populations persist in regions protected by mountainous terrain such as the Pyrenees and Caucasus. Still other priorities would be areas of high genetic diversity such as New Guinea, home to one-fifth of all the world's languages and (especially in the highlands) to an unknown fraction of the world's genetic diversity.

This wish-list begs the question of whether to sample a few dozen areas in detail or else hundreds in less detail. Here, too, the best strategy depends on the incompletely known patterns of genetic diversity that we seek to discover. For example, mtDNA sequencing detected no geographical variation within the peoples of the Middle East, but abundant variation in those of the New

Guinea highlands and among African pygmies^{2,4,5}. Similarly, the number of people to sample from each population depends on intrapopulation variability, which surely differs not only among populations but also among the genetic loci studied.

Hence one's sampling scheme will depend on the particular problem to be tackled. For instance, at one extreme mtDNA is especially variable, the control region of mtDNA is hypervariable, and the first 400 base pairs of that region contains most of its hypervariability. Of 88 mtDNA types from the control region derived from 117 Caucasian people, only 12 were shared by two or more individuals, and 10 of those 12 shared types were shared between Sardinian individuals or between Middle Eastern individuals². Again, identical mtDNA types were shared between individual !Kung or Biaka Pygmies or Mbuti Pygmies, but not between two or more of those populations⁴. At the opposite extreme, a study of nuclear DNA in Pygmies, Europeans, Chinese and Melanesians identified a polymorphism (*HP/BamHI*) whose gene frequency ranged only from 0.35 (in Chinese) to 0.46 (in Zaire pygmies)³. A larger population would be required to study variation in *HP/BamHI* than in the mtDNA control region.

By analogy, imagine trying to reconstruct the evolution of the animal kingdom if some species such as pigeons and house mice had undergone a population explosion, but if most terrestrial animals weighing over ten kilograms and most organisms inhabiting tropical rainforest had just vanished (which indeed may come to be the case before too long; the argument for sampling the genomes of such species is also compelling). The human problem is partly similar, but only partly. Although some threatened human populations may vanish because of the deaths of most members, more will vanish as their members become absorbed into other populations⁶. But attempts to sample the genomes of animals and humans have in common a need to reconcile our intellectual desire to know with our ethical desire to preserve. Above all, both share a sense of urgency, because so little time remains before it will be too late. □

Jared M. Diamond is in the Department of Physiology, University of California Medical School, Los Angeles, California 90024, USA.

1. Cavalli-Sforza, L. L., Wilson, A. C., Cantor, C. R., Cook-Deegan, R. & King, M. C. *Genomics* (in the press).
2. Di Rienzo, A. & Wilson, A. C. *Proc. natn. Acad. Sci. U.S.A.* **88**, 1597-1601 (1991).
3. Bowcock, A. M. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 839-843 (1991).
4. Vigilant, L. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **86**, 9350-9354 (1989).
5. Stoneking, M. *et al.* *Genetics* **124**, 717-733 (1990).
6. Cavalli-Sforza, L. L. *et al.* *Cold Spring Harbor Symp. quant. Biol.* **51**, 411-417 (1986).

Crystals from dreams

A. T. Winfree

RAISING the temperature of a solution in which many chemical reactions simultaneously transpire does more than just speeding up the conveyor belt of the chemical assembly line. Each reaction has its own temperature coefficient, so chemical balances also shift. This shift can be more dramatic if the supply and removal of some reactants is diffusion-limited, because diffusion coefficients are less sensitive to temperature than are typical reaction rates. Allowing reactants to mix while diffusing into a gel, Ouyang and Swinney¹ use temperature to tune the remarkable chlorite-iodide-malonic-acid reaction to create stationary chemical patterns first predicted by Alan Turing 40 years ago. Taking the temperature above and below the Turing 'bifurcation' point repeatedly erases and recreates the pattern.

The 'CIMA' reaction the authors use was first devised a decade ago^{2,3} as a substitute for the popular Belousov-Zhabotinsky (BZ) chemical oscillator. The latter is a room-temperature aqueous solution of malonic acid with bromide and bromate ions, which supports an abrupt halogen-based oxidation/reduction transition, in consequence of which the solution can be made to oscillate in colour or to propagate waves like shock-fronts of chemical transition⁴. The CIMA reaction does much the same, but something more, too, which leads us to the Turing bifurcation.

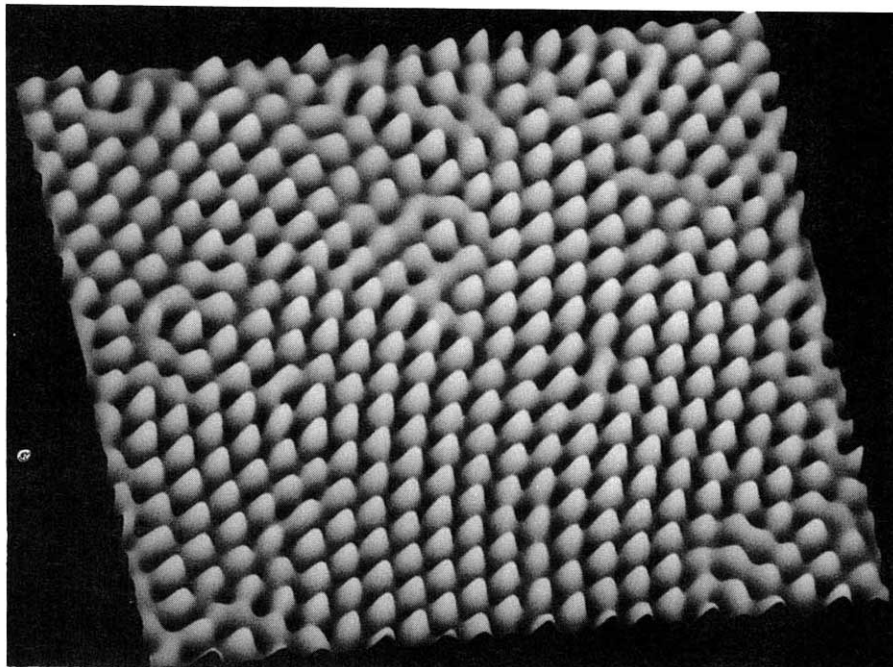
Alan Turing (associated with the Enigma Machine for cracking German codes in the Second World War, and with the advent of digital computers) introduced 40 years ago the simple but counterintuitive idea that diffusion does not necessarily increase the spatial uniformity of chemical reactions⁵. If some reactants diffuse faster than others, and if the faster species inhibit reaction while the slower ones are autocatalytic, spatial patterns may spontaneously emerge by a symmetry-breaking instability of the uniform steady-state that would persist with mixing at equal rates. Whether this happens depends, in theory, on the various reaction rates and on differences in diffusion, all of which are affected by temperature. Thus in some temperature range, a stationary pattern may spontaneously form, and in others it won't. A 'diffusive symmetry-breaking instability' or 'Turing bifurcation' is said to occur at such a transitional temperature.

Curiously, nothing of the sort was observed in the chemical laboratory, although for four decades it was the subject of countless theoretical analyses and of much speculation about biological

pattern formation. Meanwhile, the BZ reaction provided a clear laboratory example of reaction-diffusion pattern formation in the chemical 'rotor', a region of rotating chemical excitation, about 0.2 mm in radius that emits a dramatic, spiral, propagating wave⁶. Its mechanism is now well understood and the Turing bifurcation is not involved: the BZ rotor moves, does not arise

gure) which contain thousands of dots. The separation of the dots is the same 0.2 mm spacing observed since the first examples with this reagent. Were it demonstrable that the pattern really does emerge by a spontaneous symmetry-breaking instability from initial uniformity, with vigour proportional to the excess of some parameter over a threshold, it would be hard not to accept its association with the Turing bifurcation.

And this is why the report of Ouyang and Swinney is so interesting. They show that the chicken-skin pattern comes and goes and comes again as the temperature



Two-dimensional 'chicken skin' pattern made by Ouyang and Swinney with the CIMA reaction.

spontaneously from the antecedent uniform steady state, and does not require unequal diffusion.

Many were growing impatient for the discovery of a real Turing pattern when the first strong suggestion emerged in a CIMA reaction two years ago⁷. Ouyang *et al.* showed that if the ingredients of that reaction are allowed to diffuse towards one another along a one-dimensional gradient (at equal rates enhanced by convective mixing) then somewhere a colour transition occurs. With the experiment conducted two-dimensionally, this point becomes a band which breaks up laterally into dots.

This arrangement was subsequently refined⁸ to replace convective transport by strictly molecular diffusion in a gel, with similar results, which can be replicated computationally from the known reaction mechanism⁹. Further improvements¹⁰ produced up to 4 rows of dots in the gel, suggesting fully two-dimensional hexagonal arrays similar to chicken skin, with at least a hint of three-dimensional structure. Now, Ouyang and Swinney obtain two-dimensional lattices (see fi-

falls, rises and falls again through a critical 18 °C. The pattern amplitude increases with increasing coldness below that threshold. There is no hysteresis. The pattern emerges extremely slowly near threshold, as theory requires. The many dislocations in the pattern emerge independently with each erasure and renaissance, showing that they are not responses to local inhomogeneities of the substrate. If it could be shown that this two-dimensional effect vanishes when transport coefficients are strictly equalized, this would be as perfect a demonstration of Turing patterns as the

1. Ouyang, Q. & Swinney, H. L. *Nature* **352**, 610-612 (1991).
2. Orban, M., De Kepper, P., Epstein, I. R. & Kustin, K. *Nature* **292**, 816-818 (1981).
3. De Kepper, P., Epstein, I. R., Kustin, K. & Orban, M. *J. phys. Chem.* **86**, 170-171 (1982).
4. Zaikin, A. N. & Zhabotinskii, A. M. *Nature* **225**, 535-537 (1970).
5. Turing, A. M. *Phil. Trans. R. Soc.* **B327**, 37-72 (1952).
6. Winfree, A. T. *Sci. Am.* **230**(6), 82-95 (1974).
7. Ouyang, Q., Boissonade, J., Roux, J. C. & De Kepper, P. *Phys. Lett.* **A134**, 282-286 (1989).
8. Castets, V., Dulos, E., Boissonade, J. & De Kepper, P. *Phys. Rev. Lett.* **64**, 2953-2956 (1990).
9. Lengyel, I. & Epstein, I. R. *Science* **251**, 650 (1991).
10. De Kepper, P., Castets, V., Dulos, E. & Boissonade, J. *Physica D* **49**, 161-169 (1991).

most demanding theorist could require.

The actual three-dimensionality of this phenomenon remains poorly exploited. Chemical conditions must be right for Turing patterns over some range along the gradient of mixing reactants. Presumably this band moves and changes width as the temperature is changed slightly below 18 °C, enhancing the ostensible smoothness of the amplitude-temperature curve. The band's width admits at least the possibility of alternative three-dimensional structures. There is a need for invention here: a way to

maintain uniform chemical conditions throughout a thicker band.

It took almost four decades for the dreaming of theorists to accumulate a solution of Turing media so supersaturated that a real example finally crystallized only a year ago^{8,10}. More examples may be expected to precipitate now. The abundant theoretical literature suggests that many of them may be biological. □

A. T. Winfree is in the Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85718, USA.

MOTOR PROTEINS

A brave new world for dynein

Lawrence S. B. Goldstein and Ron D. Vale

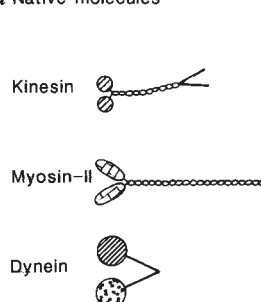
EUKARYOTIC cells rely on actin- and microtubule-based protein motors to generate intracellular movements^{1,2}. Microtubule motors come in two basic types — dyneins and kinesins — which were at first thought to move in opposite directions along the asymmetric microtubule. However, the recent finding^{3,4} that a kinesin-like motor, *ncd*, can move in the same direction as dynein along the microtubule, raised the possibility that dyneins and kinesins have similar motor domains. Surprisingly, the sequence of flagellar β -dynein heavy chain, reported on page 643 of this issue⁵ by Ogawa and on page 640 by Gibbons *et al.*⁶ shows no similarity to any kinesin-like proteins, nor indeed to any known protein.

Dynein was first discovered as a potent ATPase in cilia of the protozoan *Tetrahymena* by Gibbons in 1963. Subsequently, it was found to generate the sliding forces between the microtubule 'doublets' powering flagellar or ciliary beating. How linear displacement forces are converted into sinusoidal and other complex beat patterns is unclear, although it is known that an outer-arm dynein and several species of inner-arm dyneins contribute to this motion. Recently, another form of dynein has also been found in the cytoplasm of non-ciliated cells²; this species is implicated in chromosome movements and vesicle transport.

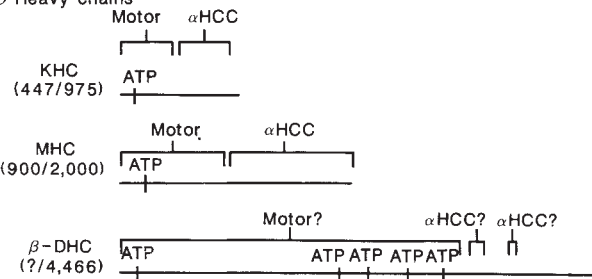
A remarkable attribute of dynein is its unusually large size. In addition to several low and intermediate-sized polypeptide chains, all dyneins contain one or more high-molecular-mass (about 500,000) heavy chains, which are the ATP-binding subunits. For example, sea-urchin flagellar dynein is a two-headed molecule containing two distinct heavy-chain subunits termed α (α -DHC) and β (β -DHC); β -DHC constitutes a functional microtubule motor independent of α -DHC^{7,8}. The total masses of flagellar and ciliary outer arm dyneins

are about 1.2×10^6 and 1.9×10^6 , respectively, a size close to that of the prokaryotic ribosome (2.5×10^6). Why is dynein so big? The large size suggests great complexity of function, although the multifaceted attributes of dynein are poorly understood in comparison to the ribosome. Some clues have emerged from biochemical and proteolytic dissec-

a Native molecules



b Heavy chains



a, Different motor molecules as seen by electron microscopy. Left, motor regions; right, α -helical coiled-coil regions of kinesin and myosin. The α - and β -subunits of dynein are indicated by the different shading of the head domains. b, Primary sequence of heavy chain components of the motor proteins shown in a. KHC, kinesin heavy chain; MHC, myosin-II heavy chain; β -DHC, β -dynein heavy chain; motor, approximate extent of known motor element; ATP, sequences that match consensus nucleotide binding sequences; α HCC, regions that may form α -helical coiled-coils; numbers in parentheses indicate size of motor and size of complete polypeptide chain in amino acids.

tions of the heavy chain, yet a better understanding has been hampered by the lack of primary sequence information.

At long last, the primary sequence of the 4,466 amino acids of a dynein heavy chain has been reported. Using two different cloning strategies and two different sea urchin species, Ogawa⁵ and Gibbons *et al.*⁶ report the complete sequence of the flagellar β -DHC. Both sets of authors have confirmed their results by independently sequencing proteolytic fragments of dynein. In addition, the sequences reported in both studies are virtually identical (97%), further strengthening the conclusions.

Several insights have emerged from the information. First, the low-resolution map of the β -DHC obtained by proteolytic digestion, ATP-vanadate cleavage

and nucleotide crosslinking has been confirmed and oriented correctly with respect to the amino and carboxy termini of the polypeptide. Second, unlike conventional kinesin and muscle myosin, analysis of the β -DHC primary sequence does not predict any region greater than 120 amino acids capable of forming an α -helical coiled-coil, which is consistent with previous data obtained from circular dichroism spectroscopy. Third, and most significant, there is no apparent sequence similarity between any portion of β -DHC and any previously sequenced motor protein, including the heavy chains of kinesins, myosins and dynamin. Thus, the β -DHC sequence seems to define a new superfamily of motor proteins.

Perhaps the most surprising feature of the β -DHC sequence, which sets it apart from kinesin and myosin, is the presence of multiple consensus sequences for nucleotide binding (see figure). One of these consensus sequences appears to belong to the primary site of ATP binding and hydrolysis. Whether the remaining four nucleotide-binding consensus sequences participate in binding ATP is not absolutely clear; however, iron- or

vanadate-mediated ultraviolet cleavage of the β -DHC in the presence of nucleotide can occur at sites other than the primary nucleotide hydrolysis region. This finding supports the contention that some of the other consensus nucleotide-binding sequences also function in ATP binding. What are the functions of these additional ATP binding sites? One tantalizing possibility is that some of these ATP binding sites participate in hydrolysis and mechanochemical reactions under particular conditions — that is, there could be multiple motor elements in one dynein heavy chain. Another possibility, pointed out by Gibbons *et al.*⁶, is that some of the ATP-binding sites could be allosteric regulators of the motor. In support of this idea, high ATP concentrations inhibit the microtubule

translocation activity of some dyneins *in vitro* (Y. Toyoshima, personal communication).

Nevertheless, the mystery of dynein's tremendous size has not been resolved by the primary sequence data^{5,6}. Although the size of the β -DHC minimal motor element is not apparent from the sequence, generating movement *per se* cannot be the reason for constructing such a large molecule, as functional motor elements of the other known motor proteins are small (450–900 amino acids). Perhaps the β -DHC motor element(s?) is also as small as its fellow motor counterparts, in which case the rest of the molecule could be a series of regulatory or attachment domains, as is true of myosins and kinesins⁹. The large size of β -DHC could also allow it to function as a mechanical spring to regulate dynein activity in response to flagellar bending.

The availability of the sequence and of clones spanning the entire length of the coding sequence establishes a foundation for elucidating how the dynein motor works. First, and most important, properties such as motor activity, microtubule binding, ATP hydrolysis and regulation can now be ascribed to specific regions of the polypeptide chain by expressing and studying defined domains. Second, regions critical for function will be revealed by sequence conservation among the next few dynein heavy chains to be cloned and sequenced. Third, analogous to the situation in the kinesin and myosin superfamilies, proteins that have dynein-like motor regions linked to other types of domains may be identified by cloning and sequencing of genes not thought to be related to dynein. Finally, it is now possible to dissect the structure–function relationships of the dynein molecule by taking advantage of the availability of dynein clones together with the recently developed methods for DNA-mediated rescue of flagellar mutants in *Chlamydomonas*¹⁰. □

Lawrence S. B. Goldstein is in the Department of Cellular and Developmental Biology, Harvard University, Cambridge, Massachusetts 02138, USA. Ron D. Vale is in the Departments of Pharmacology and Biochemistry, University of California San Francisco Medical School, San Francisco, California 94143, USA.

1. Warrick, H. M. & Spudis, J. A. *Rev. Cell Biol.* **3**, 379–421 (1987).
2. Vallee, R. B. & Shpetner, H. S. *A. Rev. Biochem.* **59**, 909–932 (1990).
3. McDonald, H. B. *et al. Cell* **63**, 1159–1165 (1990).
4. Walker, R. A. *et al. Nature* **347**, 780–782 (1990).
5. Ogawa, K. *Nature* **352**, 643–645 (1991).
6. Gibbons, I. R., Gibbons, B. H., Mocz, G. & Asai, D. J. *Nature* **352**, 640–643 (1991).
7. Vale, R. D. *et al. Cell* **59**, 915–925 (1989).
8. Sale, W. S. & Fox, L. A. *J. Cell Biol.* **107**, 1793–1797 (1988).
9. Vale, R. D. & Goldstein, L. S. B. *Cell* **60**, 883–885 (1990).
10. Mitchell, D. R. & Kang, Y. J. *Cell Biol.* **113**, 835–842 (1991).

Earthquakes in aseismic zones

PEOPLE who live near the margins of tectonic plates face the familiar but harsh reality that damaging earthquakes can disrupt their lives at any time. In contrast, people who live in the interior of continental plates do

deformation like that shown here, revealed by newly exposed, unweathered rock (John Adams is the reference height), provide valuable data on the style and extent of deformation caused by plate interior earthquakes.



Uplift of 1 m created by the Ungava earthquake of December 1989.

not consider earthquakes to be a serious threat to their lives and property. The Ungava earthquake, which occurred in northern Canada in December 1989 and is described by John Adams *et al.* on page 617 of this issue, serves as an important reminder that earthquakes are unpredictable, and especially so for those that occur in the interior of tectonic plates. The Ungava earthquake is only the tenth historical intraplate earthquake in the world to produce documented surface ruptures, so it offers earth scientists a rare opportunity to examine intraplate earthquakes using both geological and seismological data. Metre-scale

The Ungava earthquake exemplifies the challenges and obstacles faced by earth scientists who try to evaluate the hazards posed by potentially damaging intraplate earthquakes in the 'stable' interiors of continental plates. Such earthquakes occur much less frequently than those along plate margins, so seismologists have relatively few events with which to seek common traits of fault behaviour. In addition, many large intraplate earthquakes occur in areas of little or no previous seismicity, so seismological data are of limited use in identifying potentially hazardous intraplate faults. Furthermore, we know of few places in continental interiors where geologically young faulting is clearly expressed at the surface, but it is likely that our inventory is very incomplete, which makes it difficult to estimate the number of intraplate faults that might generate damaging earthquakes. The concept of plate tectonics provides a broad geological framework to explain the cause of plate-boundary earthquakes, but the factors that govern the occurrence of intraplate earthquakes are largely unknown. Seizing rare opportunities like the Ungava earthquake will permit scientists to assemble a more complete geological and seismological picture of the processes that cause crust in the 'stable' parts of continents to rupture catastrophically. (Photos courtesy of Arch Johnston.) **Anthony J. Crone**



Part of the Ungava fault viewed from a height of 10 m.

Anthony J. Crone is at the US Geological Survey, Denver, Colorado 80225, USA.

Switch to atom control

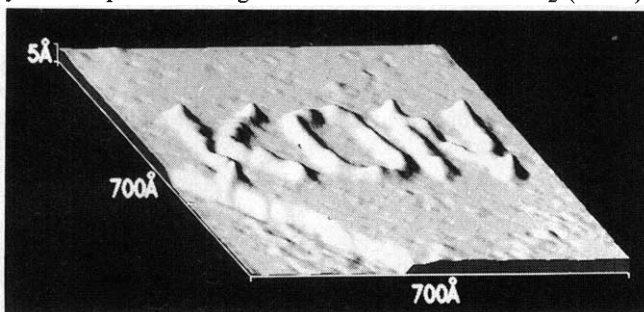
C. F. Quate

THE most dramatic evidence of the tunnelling microscope's ability to manipulate individual atoms — the construction of an atomic switch — is presented by Eigler, Lutz and Rudge on page 600 of this issue¹. This follows the previous work on translating atoms over a nickel surface using a scanning probe for guidance control². They now report on a new technique for repeatedly transferring xenon atoms back and forth between the microscope tip and substrate surface. The electrical conductance from tip to substrate depends on the position of the atom. Their device switches between a low conductance state when the atom is on the substrate to a high conductance state when the atom is on the tip. The atom switch is a bistable element and components such as these are vital in developing microcircuits.

It is not long since G. Binnig and H. Rohrer received the Nobel prize for developing the scanning tunnelling microscope, which uses a fine stylus to trace out and image surface details with atomic resolution. But these researchers themselves predicted the transition beyond a passive role for these probes when they said³ "This local probing capability [could] reach beyond its initial use in atomic-scale imaging [so that in some cases] imaging might not be used at all". This view was echoed by J. Armstrong, Chief Scientist at IBM, who said⁴ "I believe that nanoscience and nanotechnology will be central to the next epoch of the information age, and will be as revolutionary as science and technology at the micron scale have been since the early 70s".

The first report⁵ of the controlled movement of atoms with the scanning tunnelling microscope (STM) came only four years ago from Becker and colleagues of AT&T Bell Labs. They transferred a single atom of germanium from the scanning tip to the germanium substrate by suddenly increasing the top voltage to 4 volts. But techniques have developed rapidly just this year. Lyo and Avouris recently reported⁶ a more definitive experiment on a silicon surface with a precision that rivals the new atomic switch. They found that single atoms could be removed from silicon's (111) surface to the tip with a voltage pulse applied to the sample and then

redeposited on the substrate by a voltage pulse of opposite polarity. K. Mortenson and F. Besenbacher (personal communication) have used a similar approach to write lines 50 Å wide on silicon (see figure). Single atoms (presumably selenium) have also been removed in the same way with high voltages from the surface of WSe₂ (ref. 7),



Writing on silicon, using a scanning tunnelling microscope. to honour the 65th birthday of K. O. Nielsen.

as have single (sulphur) atoms from the surface of MoS₂ (ref. 8). In the latter experiment, by S. Hosoki and colleagues at Hitachi, the precision of extraction was so high that the researchers could produce lines the width of a single atom.

Whitman and colleagues have determined⁹ that the diffusion of caesium atoms over the surface of gallium arsenide is enhanced by the application of a strong electric field. The caesium atoms are attracted by the field

UNZEN VOLCANO

Anatomy of an eruption

Yoshiaki Ida

THE avalanche of hot debris that hurtled down the flanks of Mt Fugendake on 3 June this year in southern Japan marked the end of 200 years of quiescence of Unzen Volcano. The last time the volcano erupted, in 1792, 15,000 people were killed by mudslides triggered by an earthquake associated with the volcanic activity. This time, 41 were killed, mostly journalists and volcanologists¹. The fast-moving, pyroclastic flows, a trait not previously associated with Unzen, were responsible.

Unzen is set in a graben, a line of subsidence, created by north-south extension of the peninsula (see figure). Historical records of the volcano's activities report two eruptive events, in 1663 and 1792 (ref. 2). On each occasion, the eruptions started with plumes of steam and ash shooting out from the summit craters of Mt Fugendake. Lava flows with andesitic to dacitic compositions

gradient near the tip and so became concentrated beneath the tip.

The prospect opened by this remarkable body of new work is the construction of electronic devices with atomic dimensions. But this will not be so easy in practice. Speed is the limiting factor. The dilemma is illustrated with memory-type devices. For example, a cluster of 1,000 atoms could represent one bit of information, in which case the entire contents of the Library of Congress, equalling 200 terabits, could be stored on a silicon disk 12 inches in diameter. (In contrast, it would probably take 250,000 compact disks of a similar size to store this information with current techniques.) But even if one could create clusters at the exceptional rate of 10⁷ per second it would take 230 days to fill the disk. Worse still, it would take an additional 230 days to read the information. The solution will be to work towards massively parallel reading and writing systems. □

C. F. Quate is in the Edward L. Ginzton Laboratory, Stanford University, Stanford, California 94305, USA.

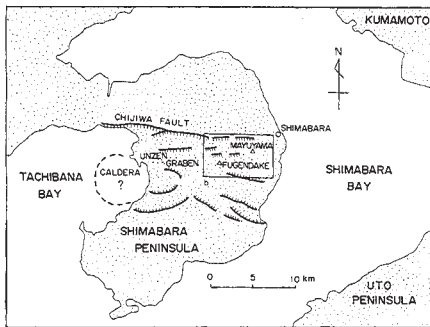
1. Eigler, D. M., Lutz, C. P. & Rudge, W. E. *Nature* **352**, 600–603 (1991).
2. Eigler, D. M. & Schweizer, E. K. *Nature* **344**, 524–526 (1990).
3. Binnig, G. & Rohrer, H. *IBM J. Res. Develop.* **30**, 355–269 (1986).
4. Armstrong, J. A. *Creativity* **10**, 2–5 (1991).
5. Becker, R. S., Golovchenko, J. A. & Swartzentruber, B. S. *Nature* **325**, 419–421 (1987).
6. Lyo, I. W. & Avouris, P. *Science* **253**, 173–176 (1991).
7. Fuchs, H. & Schimmel, T. *Adv. Mater.* **3**, 112–113 (1991).
8. *The Economist* 82 (2 February 1991).
9. Whitman, L. J., Stroscio, J. A., Dragoset, R. A. & Celotta, R. A. *Science* **251**, 1206–1210 (1991).

followed on the north flank. Neither event produced pyroclastic flows (fluidized flows of ash and hot debris). The 1792 eruption claimed as many as 15,000 lives, but this was the result of the collapse of Mt Mayuyama following the eruption: a strong earthquake triggered the collapse, causing mud to flow into the sea and generating a tsunami wave that reached the other coast of the Shimabara bay, 20 km away.

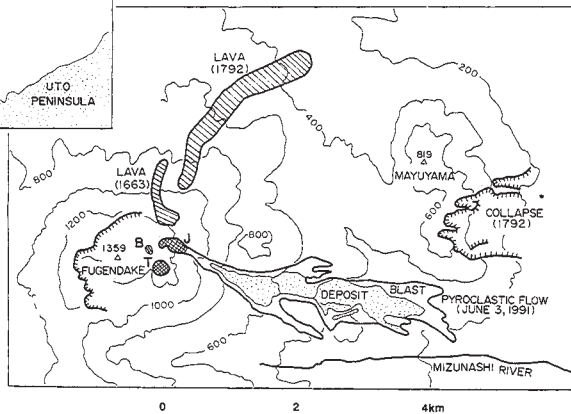
The first precursory phenomenon preceding the recent eruption was probably an earthquake swarm in November 1989 centred below a presumed old caldera (zone of collapse) in the Tachibana bay. Subsequent swarms with shallower hypocentres gradually extended the seismic area eastwards, towards Mt Fugendake, suggesting that volcanic gas and magma were moving towards the summit crater from a magma chamber below the caldera. Volcanic tremors followed in

July 1990. Finally, steam containing ash blew up from Tsukumojima and Jigokuato craters on 17 November 1990, and from Byobuiwa crater on 12 February 1991. Volcanic ash that had accumulated on the slopes produced small mudflows after heavy rain in May.

The eruptive activity entered a new stage in the middle of May. The seismic networks of the Japan Meteorological Agency and Shimabara Earthquake and Volcano Observatory detected many



Above, the tectonic setting of Unzen Volcano and, right, the eruptive activities of Mt Fugendake and Mt Mayuyama. (J, B and T are Jigokuato, Byobuiwa and Tsukumojima craters, respectively.)



shallow volcanic earthquakes immediately below the active craters. Simultaneously, a continuous electronic distance meter measurement by the Geological Survey of Japan and a tilt measurement by the Earthquake Research Institute revealed significant swelling of Mt Fugendake. On this evidence, the Coordinating Committee for the Prediction of Volcanic Eruption issued a warning of magma outflows. A 40-metre-wide lava dome appeared in Jigokuato crater on 20 May, which soon broke up and then reformed.

The major eruption of 3 June was preceded by several smaller ones, starting on 24 May, that produced pyroclastic flows in the east valley up to 2 km from the crater. The lethal pyroclastic flow of 3 June resulted from an explosion that produced a 300-metre long groove connected to the crater on the east slope. Subsequently, lava domes grew along the groove as well as on the crater. Further explosive eruptions on 8 and 11 June shot cinders up to 15-centimetres across over Shimabara city and spread volcanic ash over distances exceeding 100 kilometres. Both eruptions were accompanied by a change in tilt angle corresponding to a rapid contraction of Mt Fugendake, followed by a slower recovery.

After these events, eruptive activity

became less explosive and the domes showed evidence of lava flow in them. An estimated 0.01 cubic kilometres of material was ejected by the end of June, half the amount erupted in 1792. The magma was of a dacite composition with about 66 weight per cent of SiO_2 , according to the analysis by Kyushu University.

It is encouraging that the migration of magma to the surface was correctly predicted by the volcanologists. But it was not expected that explosive eruptions would follow the formation of lava domes. This abnormal sequence may have occurred³ because the first magma ascended into a cool environment and rose as solid lava domes. These would have prevented dissolved gas from leaking from the underlying hotter magma.

Instead, it escaped by driving the subsequent explosions. If this is the case, the explosion can be described, using the term coined in Fink's recent News and Views⁴, by a pressure-cooker model. The pyroclastic flows, Fink suggested⁵ shortly after the eruptions, arose from the collapse down the mountain flanks of the continuously growing lava domes. The collapse created a hot avalanche of exploding gas-rich lava, known as a Merapi-type flow (after the Indonesian volcano of that name). Implicit in this analysis is the prediction that the avalanches will be small-scale. Other volcanologists, however, doubt that the cooled lava domes could have the necessary explosiveness and argue instead that the outflows were Pelean or Soufriere-type, emanating directly from the volcano's vent. What actually happened will be revealed in the future, when the abundant data have been examined. □

Yoshiaki Ida is at the Earthquake Research Institute, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan.

1. Swinbanks, D. *Nature* **351**, 511 (1991).
2. Ohta, K. *Unzen Volcano - Geography, Geology and Volcanic Phenomena, Nagasaki Prefecture - 98* (Nagasaki, 1984).
3. Jaupart, C. & Allègre, C. J. *Earth planet. Sci. Lett.* **102**, 413-429 (1991).
4. Fink, J. *Nature* **352**, 188 (1991).
5. Fink, J. *Nature* **351**, 611 (1991).

Fake fur

WHEN you eat an apple, a slice of the seed-case membrane often sticks maddeningly between your teeth. How can such a frail film be driven like a nail into so narrow a gap? The penetrating seed membrane is of course supported against buckling by the surrounding apple material. Even so, says Daedalus, saliva makes the crucial contribution.

Biochemists may regard saliva as simply a solution of pre-digestive enzymes. But to an engineer, it is a cunningly formulated cutting-lubricant, designed to minimize the load on our teeth and jaws. With its subtle nonlinear visco-elasticity and low surface energy, it penetrates into food, coating the interfaces, weakening the adhesion between them, and almost abolishing the frictional drag of shearing and sliding. As an annoying side-effect, it also lubricates the insertion of food scraps between the teeth.

DREADCO engineers are now trying to imitate this effect. They are lubricating thin fibres of nylon, glass or silica with various visco-elastic fluids or waxes, and firing them at polymeric and even metallic surfaces to drive them in. They hope to perfect a lubricant wax whose rheology imitates that of saliva, and whose surface energy is matched to that of the material to be penetrated, thus minimizing the energy of insertion. A sheet of this wax will be loaded with millions of parallel fibres, packed densely together like the bristles in a brush. The sheet will be laid against a surface and flattened onto it with a sudden powerful blow. Each fibre, lubricated and protected from buckling by the surrounding wax matrix, will be driven into the surface and firmly held. When the wax is melted away, it will leave a furry surface.

So many animals have fur, says Daedalus, that it must be more than a mere camouflage and thermal insulator. His new process should bring its many benefits to technical products as well. Furry cars, for example, would absorb small collisions without damage, and would be less damaging in their turn. They would shed water as a thatched roof does, keeping the underlying metal dry; they would be quieter, and their air-resistance might be lower. They would never need painting or polishing, though they might need grooming. Furry houses would hold the heat better, and furry furniture would be softer and kinder to the touch. Even aircraft might benefit. Daedalus reckons that the soft and flexible avian wing is far more efficient than one of rigid metal alloy. Once the process has been properly worked out, he hopes to cover the wings of our current metal monsters with a delightful fledging of artificial feathers. David Jones

Riddle of the giant panda

SIR — The giant panda has been classified either into the Ursidae (bear family) based on comparative anatomical studies¹, immunological distances², DNA hybridization, isozyme electrophoresis, immunological and karyological evidence³, and palaeontological information⁴, or into the Procyonidae, based on behavioural evidence^{5,6} and haemoglobin sequences⁷. This remains a controversial issue.

We have analysed mitochondrial DNA restriction-fragment length polymorphism (RFLP) in a giant panda, a lesser panda (*Ailurus fulgens*), an Asiatic black

of mitochondrial DNA in every cell; the selection pressure on this DNA is very low. So the similarities between the two pandas, at least on RFLPs from mitochondrial DNA, may not be the result of convergent evolution. If we accept that some similarities both between the two pandas and between the giant panda and the bears are due to common descent, it is certainly a fascinating evolutionary problem to determine the cause of these similarities.

YA-PING ZHANG

LI-MING SHI

Laboratory of Cellular and Molecular Evolution,
Kunming Institute of Zoology,
Academia Sinica,
Kunming 650107, China

MITOCHONDRIAL DNA PAIRWISE
DISTANCES OF CARNIVORES

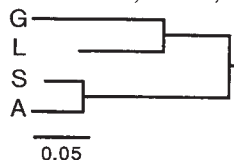
	A	B	C	D
A—Giant panda		0.177	0.326	0.327
B—Lesser panda	0.3457		0.303	0.268
C—Asiatic black bear	0.1412	0.1628		0.072
D—Sun bear	0.1409	0.2000	0.6479	

The proportion of recognition sites shared by each pair are below the diagonal.

bear (*Selenarctos thibetanus*) and a sun bear (*Helarctos malayanus*), all of which had died of illness or other accidents. We purified the DNAs from liver as described in our previous report⁸. We used 15 restriction endonucleases which recognize 6 base pairs (*Xba*I, *Bgl*II, *Eco*RI, *Eco*RV, *Pst*I, *Cla*I, *Sca*I, *Xho*I, *Sac*I, *Hae*II, *Bgl*III, *Hpa*I, *Dra*I, *Sal*I and *Bam*HI to generate restriction fragments, and observed 34–45 sites in our samples. We calculated the pairwise genetic distances (number of nucleotide substitutions per site) by the method of Nei and Li⁹ (see table). We have also constructed a molecular-phylogenetic tree for these four species of carnivora by using the neighbour-joining method¹⁰ (see figure).

In our phylogenetic tree, the giant panda is more closely related to the lesser panda than to the bears. Our results indicate that the two pandas are closely related on mitochondrial DNA RFLP. So, the key point of the controversy is whether the similarities between the two pandas are all caused by convergent evolution.

There are about 1,000–10,000 copies



Phylogenetic relationship of carnivores based on mitochondrial DNA genetic distance. Genetic distance is defined as the number of nucleotide substitutions per site. The topology was constructed using the method in ref. 10. G, giant panda; L, lesser panda; A, Asiatic black bear; S, sun bear.

in the dopaminergic nerve endings in the striatum⁶. Thus, there is reason to question the rationale of direct administration of MPP⁺ into the substantia nigra and, in our opinion, the method used by Turski *et al.* cannot be considered a model for Parkinson's disease.

Note also that Turski *et al.* cite evidence that the anticholinergic drugs used to treat Parkinson's disease are potent NMDA antagonists. These agents were the original treatment for this disease but were largely replaced by L-dopa, which is much more efficacious. In their long clinical use, anticholinergics have been demonstrated to have only acute efficacy and have no effect on the course of the disease, as implied by Turski *et al.* We write these comments because we are concerned that this study will become a basis for clinical trials with NMDA antagonists on Parkinson's disease in humans.

FRANZ HEFTI

TIMOTHY L. DENTON

KLAUS D. BECK

Andrus Gerontology Center,
University of Southern California,
Los Angeles, California 90089, USA

PATRICK P. MICHEL

INSERM U 289,
Hôpital de la Salpêtrière,
75013 Paris Cedex, France

1. Davis, D. D. *Fieldiana, Zool. Mem.* 3 (Chicago Nat. Hist. Museum, 1964).
2. Sarich, V. M. *Nature* **245**, 218–230 (1973).
3. O'Brien, S. J. *et al. Nature* **317**, 140–144 (1985).
4. Van Valen, L. M. *Nature* **317**, 428 (1985).
5. Morris, D. & Morris, R. in *The Giant Panda* 160–169 (Penguin, New York, 1981).
6. Ewer, R. F. *The Carnivora*, 393 (Cornell University Press, Ithaca, 1973).
7. Tagle, D. A. *et al. Naturwissenschaften* **73**, 511 (1986).
8. Zhang, Y. P. & Shi, L. M. *Chinese J. Genet.* **16**, 325–338 (1989).
9. Nei, M. & Li, W. H. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5269–5273 (1979).
10. Saitou, N. & Nei, M. *Molec. Biol. Evol.* **4**, 406 (1987).

No relevance to Parkinson's

SIR — Turski *et al.*¹ report that NMDA (N-methyl-D-spartate) antagonists protect neurons of the substantia nigra region of the brain from toxicity produced by local administration of MPP⁺ (1-methyl-4-phenylpyridinium, the active metabolite of the parkinsonism-inducing compound MPTP). We do not quarrel with the primary data, but we challenge the interpretation that these findings are relevant to Parkinson's disease.

Several investigations have shown that high concentrations of MPP⁺ exert dramatic, non-selective neurotoxic actions. Selective dopaminergic neurotoxicity, the key characteristic of Parkinson's disease, occurs only when dopaminergic neurons are exposed to a limited range of MPP⁺ concentrations for example, in cultures of mesencephalic neurons, 0.1–30 μ M MPP⁺ selectively destroys dopaminergic neurons, whereas all cells are destroyed at higher concentrations². Turski *et al.* infused a 25 mM MPP⁺ solution into the substantia nigra. But such concentrations of MPP⁺ indiscriminately destroy all cells at the site of injection (refs 3–5; our unpublished observations). Furthermore, MPP⁺ does not accumulate significantly in nigral cell bodies, but rather

1. Turski, L., Bressler, K., Rettig, K. J., Loschmann, P. A. & Wachtel, H. *Nature* **349**, 414–418 (1991).
2. Michel, P. P., Dandapani, B. K., Sanchez-Ramos, J., Elfange, S., Pressman, B. C. & Hefti, F. J. *pharmac. exp. Ther.* **285**, 842–850 (1989).
3. Heikkilä, R. E., Nicklas, W. J. & Duvoisin, R. C. *Neurosci. Lett.* **58**, d135–140 (1985).
4. Altar, C. A., Heikkilä, R. E., Manzini, L. & Marien, M. R. *Eur. J. Pharmac.* **131**, 199–209 (1986).
5. Gibb, W. R. G., Costall, B., Domeney, A. M., Kelly, M. E. & Naylor, R. J. *Brain Res.* **461**, 361–366 (1988).
6. Herkenham, M. *et al. Neuroscience* **40**, 133–158 (1991).

A planet, not a plasma cloud?

SIR — Helfand and Hamilton¹ have suggested that the periodic delay and advance in arrival times of signals from the pulsar PSR1829–10, which we attributed to the existence of a planet orbiting the pulsar², could instead be caused by a stationary dispersing cloud of material, about 1 AU in size, quite close to the Sun on the line of sight to the pulsar. This would naturally account for the 6-month variation in pulse arrival times but, as I understand their model, when the Earth–Sun–pulsar angle is 90° the pulse arrival-time residual should be a minimum (that is, when we are seeing 'around' the dispersive cloud, the signal transmission time is shortest). This is not what is observed: as Helfand and Hamilton correctly state, at 90° (for example, at MJD 48166), we observe nearly zero residual, not a minimum. The nearest minimum is about 1.5 months away, so

that the coincidence in phase relationship between the Earth's orbit around the Sun and that of the planet around the pulsar is not explained by the model of Helfand and Hamilton, but rather argues against it.

There is further relevant observational evidence, concerning the frequency dependence of the amplitude of the sinusoid formed by the timing residuals. The data set used in ref. 2 contains observations at frequencies around 1,408 MHz and 1,630 MHz. Independent fits at these frequencies give amplitudes of 7.7 ± 0.3 ms and 7.4 ± 0.6 ms respectively, whose ratio is 1.04 ± 0.08 . This is significantly different from the value of 1.34 that would be expected if the dispersing cloud were a cold plasma, because the magnitude of any delay will depend on the inverse square of the observing frequency. A ratio of 1 is of course what

one would expect for the planetary interpretation, in the absence of any dispersing clouds.

Two other aspects of this model also worry me. First, it seems most unlikely that the structure of the cloud should give rise to an integrated density profile that produces such a closely sinusoidal form in the observed time delay. Second, such a large, dense plasma cloud so close to the Sun would be expected to produce substantial H α emission, which to my knowledge is not observed.

ANDREW LYNE

*University of Manchester,
Nuffield Radio Astronomy Laboratories,
Jodrell Bank,
Macclesfield, Cheshire SK11 9DL, UK*

1. Helfand, D. J. & Hamilton, T. T. *Nature* **352**, 481 (1991).
2. Bailes, M., Lyne, A. G. & Shemar, S. L. *Nature* **352**, 311–313 (1991).

Another look at the Big Bang

SIR — In a discussion¹ on the resolution of Olbers' paradox, prompted by Wesson's article², Maddox compares the effects of the age and expansion of the Universe. The age limits the light emitted by stars, and expansion redshifts the starlight. The age limit was first calculated by Lord Kelvin³ in 1901, who showed that the night sky is dark because the visible Universe, having a size determined by the finite speed of light, contains too few stars to cover the whole sky. Bondi in 1952 attributed darkness⁴ of the night sky to expansion and argued that the darkness is direct proof of the expansion of the Universe. Bondi's popular argument, true in a steady-state universe of infinite age, fails in an evolving universe of finite age. For, if Kelvin's solution applies in a static universe uniformly populated with stars, then it also applies in an expanding universe of galaxies, and expansion, absorption, and clustering of stars serve only to make the night sky darker.

I discussed⁵ the age limit in 1964, and showed that to fill a static universe with starlight in thermodynamic equilibrium with stars requires that stars shine continuously for 10^{23} years. The energy density of starlight in a static universe only 10^{10} years old is 10^{-13} of that needed to create the bright sky of Olbers' paradox. Even in a much older universe, the sky remains dark because the lifetime of luminous stars is typically 10^{10} years.

I also showed⁵ that the effect of expansion is usually small. Two universes of equal age, one static and the other expanding, with stars in identical states, contain equal numbers of photons emitted by stars. In the expanding universe the average redshift of photons is $\langle z \rangle = (1 + q)^{-1}$ and the radiation energy

density is $u_s / (1 + \langle z \rangle)$, or $u = u_s (1 + q) / (2 + q)$, where u_s is the radiation energy density in the static universes, and the deceleration term q is constant. In all decelerating universes of $q > 0$, expansion reduces the radiation to a level never less than 0.5 of that in the static universe. Thus, in the standard model (density parameter equal to unity), $q = 0.5$, hence $u = 0.6u_s$. Bondi's redshift solution, applied to a Big Bang universe, reduces the radiation field by a factor not less than 0.5, whereas Kelvin's solution limits the radiation field to 10^{-13} of that needed to create a bright sky. The age effect clearly dominates.

The visible Universe has a radius roughly 10^{10} light years. Stars in a sphere of radius 10^{10} light years cover geometrically (not diffractively) 10^{-13} of the sky. Insufficient stars exist in the visible universe to cover the sky. In the long history of the paradox⁶, Kelvin was the only person to relate the radiation density of starlight to the sky-cover fraction⁷, thereby proving the aptness of Olbers' geometric line-of-sight argument⁸.

When the effect of expansion dominates, as in Bondi's steady-state solution, the sky is covered with stellar disks, most of which cannot be seen because of large redshift. When the effect of age dominates, as in Kelvin's solution, the sky is not covered with stellar disks; stars in the visible Universe must increase in number by 10^{13} to cover the sky and come into equilibrium with starlight.

Olbers' paradox is full of intricacy and subtlety, which explains why it continues to fascinate, if only for theoretical reasons. Substitution of galaxies for stars changes the name of the game: a more suitable name would be 'Bonnor's paradox'⁹, were it not that a sky tiled with galaxies is not in paradoxical con-

flict with observation.

Olbers' paradox has climaxed this century with the realization that the sky is covered not by redshifted stars, but by the redshifted Big Bang. At night we look out between the stars to immense distances and far back in time before the formation of galaxies and stars. We look out and back to the limit of the visible Universe, and almost every line of sight terminates in the early Universe. The Big Bang covers the entire sky, but its brilliance is mercifully redshifted by the expansion of the Universe into an infrared gloom invisible to the naked eye.

EDWARD HARRISON

*Physics and Astronomy Department,
University of Massachusetts,
Amherst,
Massachusetts 01002, USA*

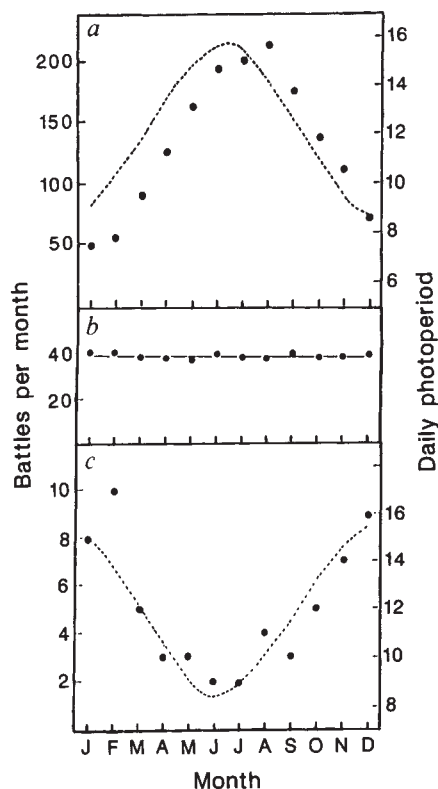
1. Maddox, J. *Nature* **349**, 363 (1991).
2. Wesson, P. S. *Astrophys. J.* **367**, 399–406 (1991).
3. Thomson, W. *Phil. Mag. Ser. 6*, **2**, 161–177 (1901).
4. Bondi, H. *Cosmology* (Cambridge University Press, 1952).
5. Harrison, E. R. *Nature* **204**, 271–272 (1964).
6. Harrison, E. R. *Darkness at Night: A Riddle of the Universe* (Harvard University Press, Cambridge, Massachusetts, 1987).
7. Harrison, E. R. *Nature* **322**, 417–418 (1986).
8. Olbers, W. H. *Astronomisches Jahrbuch 1826* (ed. Bode, J. E.) (Späthen, Berlin, 1823). (Trans. in ref. 6.)
9. Bonnor, W. *Mon. Not. R. astr. Soc.* **128**, 33–47 (1964).

Rhythms of war

SIR — We decided to examine whether seasonal rhythms exist in the opening dates of wars. Our analysis is based on a total of 2,131 acts of hostility, appearing as entries in Laffin's monograph¹, for which a defined opening month date is given. Each act of hostility was assigned both its opening month date and its line of latitude.

In the Northern Hemisphere, latitudes 30–60° N, the annual rhythm in the opening dates of wars shows a peak in August and a nadir in January (*a* in the figure). An inverse pattern in the annual rhythm of wars with a peak in December–February and a nadir in July was found in the Southern Hemisphere, latitudes 30–60° S (*c* in the figure). The circannual rhythms in the acts of aggression, both in the Northern and the Southern hemispheres, have a statistically significant positive correlation with the annual rhythm of the photoperiod duration in the same geographical regions. The results in the Northern Hemisphere suggest that there is a phase-shift of about one month between the two rhythms. We found a constant rate of acts of hostility throughout the year around the line of the Equator (*b* in the figure).

Barbara Tuchman² designates affectiveness and irrationality as dominant contributory factors in the series of wars from Troy to Vietnam. Elongation of the daily photoperiod may induce increases



The monthly rate of the opening dates of wars at latitudes 30°–60° N (a) and 30°–60° S (●) are correlated with the duration of the daily photoperiod. Data for the photoperiod duration at latitude 45° N (a) and 45° S (c) (---) are taken from ref. 4. The correlation between the monthly rate of wars in (a), and the duration of the daily photoperiod were found to be highly statistically significant. Spearman's rank coefficients 0.834, $P < 0.001$ and 0.86, $P < 0.001$, respectively. A phase-shift of one month in advance in the photoperiod rhythm resulted in improved correlations: Spearman's rank coefficients 0.991, $P < 0.001$ and 0.986, $P < 0.001$, respectively. the correlation between the monthly rate of wars in (c) and the duration of the daily photoperiod was found to be highly statistically significant. Spearman's rank coefficient 0.912, $P < 0.001$. A phase-shift of one month in advance in the photoperiod rhythm did not result in improved correlation. Opening dates of wars around the line of Equator (latitudes 30° N–30° S (b) show a constant monthly rate throughout the year.

in affective aggressiveness which results in the correlation described here. Could the photoperiod environmental cue be one of the factors which call the "lust for hatred and destruction"³ into play?

GABRIEL SCHREIBER
SOFIA AVISSAR
ZEEV TZAHOR
NIMROD GRISARU

Beer-Sheva Mental Health Centre and
Ben Gurion University,
PO Box 4600,
Beer-Sheva, Israel

1. Laffin, J. *Brassey's Battles: 3500 Years of Conflict* (Pergamon, London, 1986).
2. Tuchman, B.W. *The March of Folly: from Troy to Vietnam* (Knops, New York, 1984).
3. Einstein, A. *The Complete Psychological Works of Sigmund Freud*, 22, 199–202 (Hogarth, London, 1932).
4. *The Times Atlas of the World* (Bartholomew, London, 1967).

Release factors and genetic code

SIR — Meyer *et al.* reported¹ that UGA was a code for the amino acid cysteine in the ciliate *Euplotes octacarinatus*, in addition to the 'universal' cysteine codons UGU and UGC. UAA was its stop codon, and codon UAG was not found^{1,2}. Meyer *et al.* suggested that the "role of the classic termination codons had not yet been established when the ciliates started to diverge from other eukaryotes". We have already proposed³ a different explanation: divergent codes in ciliates, such as *Tetrahymena*, "may be formed by capture of stop codons". Similarly, Miceli *et al.*², referring to *Euplotes*, suggested that "the eccentric genetic code of some ciliates is a derived and not a primitive character".

Release factors in eukaryotes normally recognize stop codons UAA, UGA and UAG. In *Euplotes*, these factors appear to be specific for UAA^{1,2,4}, and, if so, codons UGA and UAG would become untranslatable nonsense codons which disappear from coding sequences⁵. We suggest that UGA reappeared as a cysteine codon when G in cysteine anticodon GCA became modified, perhaps to I (inosine) so as to pair with U, C and A in the third positions of codons, UGU, UGC and UGA. Similar modifications of G in anticodons, to pair with U, C and A, have been proposed for anticodons GUU (asparagine) and GAU (isoleucine) in echinoderm mitochondria⁶, and for anticodon GUA (tyrosine) (in addition to anticodons GUU and GAU) in planarian mitochondria (Y. B. *et al.*, manuscript submitted).

In *Tetrahymena*, release factors became specific for UGA (rather than UAA) so that UAA and UAG became untranslatable and disappeared. These codons would have been captured by glutamine³ when two additional glutamine anticodons were formed by mutation of anticodon UmUG to UmUA in a duplicate of glutamine transfer RNA, followed by mutation of a duplicate of codon UmUA to CUA⁷. UAA and UAG were translated as glutamine in *Tetrahymena* and the ciliates *Oxytricha*, *Paramecium* and *Stylonychia* when they reappeared, by mutations of some CAA and CAG glutamine codons³. An alga, *Acetabularia*, presumably followed the same course of events⁸.

Planarian mitochondria evidently use UAG as the only stop codon. UGA encodes tryptophan and UAA (in addition to UAU and UAC) for tyrosine. One release factor, RF2, which recognizes stop codons UGA and UAA in prokaryotes, has not been detected in rat mitochondria⁹. Presumably it has been deleted, because it would read UGA as

stop¹⁰. Another release factor, RF1, which normally recognizes stop codons UAA and UAG, has apparently become specific for UAG in planarian mitochondria, making UAA untranslatable so that it disappeared. UAA was captured by tyrosine when anticodon GUA became modified, presumably to IUA, and UAA reappeared from mutations of tyrosine codons UAU and UAC (Y.B. *et al.*, manuscript submitted; see ref. 10). This change seems analogous to the one with UGA in *Euplotes*.

The capture of UGA by cysteine in *Euplotes* probably occurred after the divergence of *Euplotes*¹¹ as a recent event in ciliate phylogeny. The changes we have suggested accord with the codon capture theory for non-disruptive change in the genetic code⁵.

THOMAS H. JUKES

Space Sciences Laboratory,
University of California, Berkeley,
6701 San Pablo Avenue,
Oakland, California 94608, USA

YOSHITAKA BESSHO

TAKASHI OHAMA

SYOZO OSAWA

Department of Biology,
Nagoya University,
Chikusa-ku, Nagoya 464-1, Japan

1. Meyer, F. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **88**, 3758–3761 (1991).
2. Miceli, C. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **86**, 3016–3020 (1989).
3. Jukes, T. H., Osawa, S. & Muto, A. *Nature* **325**, 668 (1987).
4. Harper, D. S. K. & Jahn, C. L. *Proc. natn. Acad. Sci. U.S.A.* **86**, 3252–3256 (1989).
5. Osawa, S. *et al.* *Proc. R. Soc. B* **241**, 19–28 (1990).
6. Himeno, H. T. *et al.* *Gene* **56**, 219–230 (1987).
7. Hanyu, K. *et al.* *EMBO J.* **5**, 1307–1311 (1986).
8. Schneider, S. U. *et al.* *Molec. gen. Genet.* **218**, 445–452 (1989).
9. Lee, C. C. *et al.* *J. biol. Chem.* **262**, 3548–3552 (1987).
10. Jukes, T. H. & Osawa, S. *Experientia* **46**, 1117–1126 (1990).
11. Lynn, D. H. & Sogin, M. L. *BioSystems* **21**, 249–254 (1988).

Ice age carbon

SIR — Two recent attempts have been made to understand past changes in the global carbon cycle by estimating changes in carbon storage during the ice age. One was based on a climate-biosphere model¹, the other on palaeoecological inferences². Changes in carbon storage are thought to be related to carbon dioxide levels during the last glacial maximum, 18,000 years before present (BP). The two approaches give very different estimates for change in terrestrial carbon storage: ± 50 giga-tonnes (10^9 g) for the former¹ and about 1,350 Gt for the latter². But the marine record supports neither conclusion.

Marine $\delta^{13}\text{C}$ data can be used to constrain terrestrial-based estimates of carbon variations, as some of the marine changes in the glacial-interglacial reservoir reflect effects due to the transfer of

carbon from land to ocean³. One advantage of the palaeoceanographic approach is that it can provide a robust global estimate of variations on land. Using the best marine estimate of the terrestrial reservoir effect at the last glacial maximum⁴, and the standard relation between carbon isotope variations and terrestrial organic carbon³, I calculate that about 450 Gt were transferred from land to the ocean. Thus, the marine record supports the conclusions neither of ref. 1 nor ref. 2, and indicates a significant discrepancy in our understanding of carbon transfer between the two reservoirs.

The magnitude of the marine-terrestrial discrepancy is not trivial — nearly 1×10^{12} tonnes of carbon in the case of ref. 2. And if one considers only the area of land covered by ice and tundra at the last glacial maximum^{2,5}, the glacial-interglacial carbon changes in these areas alone were sufficient to account for most of the marine variations.

Using the above high-latitude terrestrial changes as a constraint on reconciling marine and palaeoecological approaches, three possible explanations are: (1) terrestrial carbon changes in the tropics were significantly less than estimated in ref. 2; (2) a large amount of carbon was sequestered in 'marginal' reservoirs of the continental shelves and slopes; or (3) the abundance of C_4 plants (for example, many grasses), which have a significantly different isotope signature than the currently dominate C_3 plants, was significantly greater during the last glacial maximum.

There are arguments for all three of these possibilities. Evidence for lowland tropical climate change in the Amazon basin and Indonesian rainforests is weak and sometimes ambiguous⁶. There is also evidence for significant carbon sequestering on some continental shelves⁷. Lower CO_2 and greater aridity in the ice age should also have enhanced abundance of C_4 plants, although no estimates are available of the potential magnitude. Whatever the explanation, the marine record points to a significant gap in our understanding of ice-age terrestrial carbon storage and carbon exchange between marine and terrestrial reservoirs.

THOMAS J. CROWLEY

Applied Research Corporation,
305 Arguello Drive,
College Station, Texas 77840, USA

1. Prentice, K. C. & Fung, I. Y. *Nature* **346**, 48–51 (1990).
2. Adams, J. M., Faure, H., Faure-Denard, L., McGlade, J. M. & Woodward, F. I. *Nature* **348**, 711–714 (1990).
3. Shackleton, N. J. in *The Fate of Fossil Fuel CO_2 in the Oceans* (eds Andersen, N. R. & Malahoff, A.) 401–428 (Plenum, New York, 1977).
4. Duplessy, J. C. et al. *Palaeoceanography* **3**, 343 (1988).
5. Schlesinger, W. H. *Nature* **348**, 232–234 (1990).
6. Crowley, T. J. & North, G. R. *Paleoclimatology* (Oxford University Press, New York, 1991).
7. Klinger, L. F. in *Scientists on Gala* (eds Schneider, S. H. & Boston, P. J.) (MIT, Cambridge, in the press).

Amphipod invasion on the Rhine

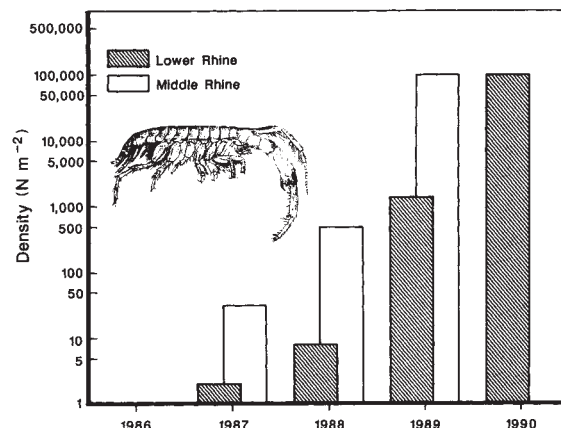
SIR — The river Rhine is a dramatic example of an ecosystem under environmental stress. River engineering, the discharge of industrial, agricultural and domestic sewage, as well as industrial cooling-water effluents have resulted in an impoverishment of the original fauna. In contrast to the decline of the native macrozoobenthos, the number of invader species in the Rhine has increased to fill in the empty niches¹. Most of these invaders are euryhaline, thermophilous and/or *r*-strategists. A recent immigrant is the amphipod *Corophium curvispinum*, which originates from the Ponto-Caspian area². Since its first colonization in 1987, its numbers have increased explosively in the middle and lower Rhine (see figure). Maximum densities of 100,000 specimens per square metre have been found, which is much greater than the numbers recorded in other rivers³. The explosion occurred first in the middle Rhine and a year later in the lower over a distance of about 200–500 km, indicating an enormous range extension (see figure).

Why is this species so successful in the

this ecosystem is immense. Because of its high breeding capacity and its small size (adults are 2.5–7.0 mm), generation time is short, which means an enormous increase in filter-feeding activity in the river. So, competition for food with other filter-feeders, such as the alien zebra mussel *Dreissena polymorpha*, the native caddis fly *Hydropsyche contubernalis* and zooplankton species, might be expected. Indeed, we found planktonic diatoms and green algae in intestinal tracts of *C. curvispinum*.

The high densities of *Corophium* could present a problem for the (re)settlement of riverine species into the Rhine ecosystem. Because the immigrant builds dense colonies on stones of breakwaters and river banks, there is a severe competition with other zoobenthos species for space, as these stones are the only available stable substrate in the Rhine. We have observed specimens of *Dreissena* which were completely overgrown by the muddy tubes of *Corophium*. The overgrowth of stones by these muddy tubes prevents settlement of *Dreissena* larvae, because these

Maximum densities of *C. curvispinum* in the German middle Rhine (open boxes; from ref. 5) and in the Dutch lower Rhine at Lobith, near the Dutch-German border (hatched boxes; own data).



Rhine? As it has a southern origin, it probably takes advantage of the increased water temperature of the river. Because it is originally a brackish-water (less than 6 per mil salinity) species⁴, *C. curvispinum* tolerates the high salinities (2.5–7.0 mM Na^+) of the Rhine, which result from mining activities in the drainage basin. Its breeding behaviour also contributes to its successful dispersal. We found that the species breeds from April until September, producing three generations a year, one more than the related species *C. volutator*, *C. insidiosum* and *C. arenarium*. These amphipods live in muddy tubes, which could provide shelter against predators. As one of the world's most eutrophicated and organically polluted rivers, the Rhine constitutes an abundant food supply for *C. curvispinum*, which is a filter-feeder³.

The impact of the recent invader on

larvae need bare substrates for attachment. *Dreissena* is an important food for diving ducks, commercial fish and crayfish. A decline in the populations of *Dreissena* from the lower Rhine has recently been reported by fishermen.

FRED W.B. VAN DEN BRINK

GERARD VAN DER VELDE

Laboratory of Aquatic Ecology,
Catholic University of Nijmegen,
NL-6525 ED Nijmegen, The Netherlands
ABRAHAM BIJ DE VAATE
Ministry of Transport and Public Works,
Institute for Inland Water Management
and Waste Water Treatment,
NL-8200 AA Lelystad, The Netherlands

1. Van den Brink, F.W.B., Van der Velde, G. & Cazemier, W.G. *Limnologie aktuell* **1**, 192–216 (1990).
2. Van den Brink, F.W.B., Van der Velde, G. & Bij de Vaate, A. *Bull. Zool. Mus. A'dam* **11**, 211–213 (1989).
3. D'Udekem d'Acoz, C. & Stroot, P. *Ann. Soc. R. Zool., Belg.* **118**, 171–175 (1988).
4. Taylor, P.M. & Harris, R.R. *J. comp. Physiol. B* **156**, 323–329 (1986).
5. Schöll, F. *Lauterbornia* **5**, 67–70 (1990).

First man of the Earth

Stephen Jay Gould

Principles of Geology, First Edition, Volumes I–III. By Charles Lyell. University of Chicago Press: 1990 (Vol. I), 1991 (Vols II and III).*

LYELL'S *Principles of Geology* (published in three volumes between 1830 and 1833) has often been viewed as the first great English textbook of the Earth. Lyell himself treated the book as modern authors use their texts: as a lifetime source of revision and income — for he issued eleven editions, the last in 1872 just three years before his death. Text books have a well-deserved reputation for insipidity engendered by a style that blots out all vigour of prose and personal belief in presenting a supposedly objective compendium of established knowledge.

If Lyell's *Principles* is a textbook (in this usual sense of terminal dullness), then why does it continue to exert such an influence more than 150 years after its publication, and why do we rejoice so heartily that the University of Chicago Press has finally issued an affordable facsimile of the nearly unobtainable first edition? (The new version is, by the way, a lovely effort graced with style in production and a fine preface by the greatest modern historian of English geology, Martin Rudwick.) The answer to this conundrum is easily given: *Principles* is not a textbook in any legitimate sense of the term, either then or now. The book is a long brief for a definite (and in many ways idiosyncratic) point of view, not only on the operation and history of the Earth, but about proper procedure in all of science. Lyell was a barrister by original profession, and a brilliant prose stylist by gift and practice. No scientist has ever made a partisan argument more forcefully and effectively.

Lyell certainly understood his enterprise as a battle for an unconventional point of view. In a private letter, he wrote about his intent in *Principles*: "I am grappling not with the ordinary arm of flesh, but with principalities and powers, with Sedgwick, Whewell and others, for my rules of philosophising, as contradistinguished from them, and I must put on all my armour." When we recall that, in the full citation (Ephesians 6:12–13),

Paul includes among the enemies "rulers of the darkness" and "spiritual wickedness in high places", and that he donned his armour to "be able to withstand in the evil day", we get a better sense of Lyell's self-perceived mission.

Lyell's most famous disciple, Charles Darwin, wrote of his own much shorter book *The Origin of Species*: "This whole



Lyell's insistence on a nondirectional world allowed him to speculate that extinct forms might reappear to grace our Earth anew. The caption to this satirical cartoon, drawn by one of Lyell's colleagues in response to this vision, reads: "You will at once perceive," continued Professor Ichthyosaurus, "that the skull before us belonged to some of the lower order of animals; the teeth are very insignificant, the power of the jaws trifling, and altogether it seems wonderful how the creature could have procured food."

volume is one long argument." Lyell might have made the same claim with even more justice; for whereas scholars are still debating what Darwin meant (was he arguing for evolution itself, for natural selection as a mechanism, for a method of historical reconstruction?), Lyell stated his brief for *Principles of Geology* with admirable clarity in a letter that he wrote to his field partner (and notable fellow geologist) Roderick Impey Murchinson:

My work . . . will not pretend to give even an abstract of all that is known in geology, but it will endeavour to establish the principles of reasoning in the science . . . which, as you know, are neither more nor less than that no causes whatever have from the earliest time to which we can look back, to the present, ever acted, but those now acting; and that they never acted with different degrees of energy from that which they now exert.

This radical uniformitarianism implied a set of methodological reforms, salutary and enduring for the most part: work with observable causes extrapolated through the immensity of geological time; do not speculate about past changes outside current boundaries. But Lyell's vision extended well beyond methods of research to a set of substantive claims about the nature and history of the Earth. If the rates and ranges of current causes encompass all past changes, then large results can arise only by extended summation of small effects, for even the biggest modern earthquakes and volcanic eruptions do not convulse the entire Earth. This substantive belief in gradualism shaped and restricted

the subsequent history of geological thought (for Lyell's 'reforms' could be blinkering as well as expansive). The almost visceral disgust that is still felt by many geologists and palaeontologists for such eminently plausible (and empirically supported) notions as impact-induced mass extinctions represents the most unfortunate aspect of Lyell's legacy. In any case, we must read Lyell to grasp both his gifts and his imposed limitations.

Even more central to Lyell's vision was his highly idiosyncratic conviction that the Earth had remained in steady state throughout its history — always changing to be sure, but never in any particular direction. Seas move in and out; continents emerge and erode away; temperatures rise and fall. Lyell even argued that life had maintained an un-

changing mean complexity through time (while species continually arose and disappeared), and predicted that mammals would be found in Palaeozoic rocks. This 'uniformity of state' provides the organizing principle for Lyell's 'long argument' spanning all three volumes: the first on rates and ranges of current causes; the second on the steady state of life's forms and numbers; the third on a statistical method for geological correlation in a nondirectional world. (Lyell computed the percentage of species still living and named the epochs of the Tertiary — Eocene, Miocene and Pliocene in his original version —

*Vol. I: Introduction by Martin J. S. Rudwick; pp. 584; hbk \$39.95, £31.95; pbk \$17.95, £14.25. Vol. II: Pp. 330; hbk \$35, £27.95; pbk \$15.95, £12.75. Vol. III: Cumulative bibliography compiled by Martin J. S. Rudwick; pp. 580; hbk \$39.95, £31.95; pbk \$17.95, £14.25. (Vol. III to be published in the United States in October and in the United Kingdom in December.)

according to the increasing number of extant forms. In a nondirectional world, you cannot look for distinctive evolutionary stages in progressive sequences, and can only calculate increasing degrees of likeness to modern faunas, as old species die and new species of equal complexity emerge.)

Lyell had a vision, but he was not a dogmatist. He never abandoned the doctrine of uniformity of rate (gradualism), but he did alter his insistence upon a nondirectional Earth when the evidence of early complexity failed to materialize. Finally, he supported evolution in later editions (although he never accepted Darwin's mechanism of natural selection) — not with the joy of discovery, but with great reluctance as a minimal retreat from his preferences in the face of new evidence. (Evolution at least permitted continued adherence to gradualism when the steady state of anatomical complexity had to be dropped.) Thus, the later, post-Darwin editions of *Principles* are an entirely different book. We must read the first edition to understand Lyell's vision.

Darwin was in the process of writing a much longer book when Wallace's independent discovery of natural selection spurred him to dash off the hurried 'abstract' that we know as *The Origin of Species*. Scholars have calculated that Darwin's original plan, if executed, would have yielded a book about as long as *Principles of Geology*. So how would biology have differed if Darwin had executed his original plan? A cynical wag might answer: "Not much at all since the world was ready for evolution, except that the same number of people would have been convinced, with many fewer having actually read the book."

How many people read all of Lyell, and why should we read him today? Does a facsimile of the first edition hold more than antiquarian value? Such questions have many answers, all in Lyell's favour. Consider just two. First, Lyell is a joy to read. He was a great, not merely a good, prose stylist. We must lament the disciplinary boundaries that place scientists outside the realm of literature, for at least three nineteenth-century British scientists, John Playfair, Thomas Henry Huxley and Charles Lyell, wrote as beautifully as many of the most famous Victorian novelists; yet they never seem to get proper credit. Lyell's success owes as much to his powerful style as to the cogency of his logic or the force of his evidence. People did read Lyell, and his gorgeous prose is no less compelling today.

Second — and I don't know how to say this without sounding peremptory or condescending — we have responsibilities as scholars, particularly to uphold some ethical vision of excellence in a

largely anti-intellectual world. Our academic lineages contain a few documents so central to their later history that we cannot consider ourselves educated unless we study them. Lyell's *Principles* is the pre-eminent document for any student of the Earth. Some classics, I will admit, can be chore, whatever the necessity. But Lyell is a joy, so why fuss? To quote Paul again, such reading can only be a "labour of love" (Thessalonians 1:3), for "the first man is of the earth, earthy" (1 Corinthians 15:47). □

Stephen Jay Gould is in the Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138, USA.

Casting pearls

Walter Gratzer

The Thermodynamics of Pizza: Essays on Science and Everyday Life. By Harold J. Morowitz. Rutgers University Press: 1991. Pp. 247. \$21.95.

Basic Nature. By Andrew Scott. Blackwell: 1991. Pp.192. £16.95, \$17.95.

"THE three bases on the DNA are known as a condom", according at least to a candidate in a national examination. Well it was only a schoolboy and a sharp cuff about the ear should have ensured a better grasp of the facts. But what about the pronouncement only a few weeks ago by a member of the British intelligentsia, alumnus of an ancient university, and a minister of the crown (formerly of education, no less): speaking with eloquence and authority about the difficulties of legislating for the suppression of dangerous breeds of dogs, he declared that pit-bull terriers were cross-breeds and therefore not easily defined in law — and dogs, he was reliably informed, had no DNA.

How comes it then that so few members of the bookish classes have managed to assimilate even the rudiments of the stock of human knowledge? Harold Morowitz, reflecting on Henry David Thoreau's interest in science, cites an allusion in *Walden* to Liebig's work on metabolic energy conversion. This was then a topic at the forefront of the new discipline of biological chemistry. "How", Morowitz muses, "did the news travel so fast from Giessen to Concord?" In today's newspapers DNA appears almost as frequently as extrapolations and parameters: it has infiltrated medicine, agriculture, the law and the turf. "But what is DNA?" an intelligent, literate and well-informed friend asked me the other day. I vouchsafed that it was a molecule. "Then is it", my friend persisted, "like a colour or a smell?" We had, I decided, more arduous spade-

work before us than I was willing to contemplate.

I offer this as an example to Andrew Scott of what he is up against, for it is precisely at the sector of the population to which my friend belongs (not to mention Her Majesty's minister) that this book is directed. In fewer than 200 pages, Scott has attempted to survey the state of knowledge in science — the whole lot, from soup to savoury. He has wrought heroically in a now rather unfashionable style, for his book is austere produced, with no pictures or gimmickry, no anecdotes or purple passages. Like Stephen Hawking, he has permitted himself but one equation, and it is the same one. Scott's *tour d'horizon* sets out from 'Spacetime' and ends with 'Brains' and 'Mystery'. The writing is unfailingly lucid, although I wonder how the reader who needs to have it explained that 10^{-27} is how grown-ups represent 0.000...1 will fare with the dissertation on the properties of clocks in space.

Scott is a chemist and he comes on strongly on reaction mechanisms, but there is practically no mention of organic chemistry, nor does it emerge that the chemistry of life is that of carbon; and for all Scott's obvious fascination with the origins of life, the nature of water is not discussed. (Even the Pope, J. B. S. Haldane once observed, is 70 per cent water.) But these are merely cavils; Scott can be read with profit by everyone from ministers of the crown on up.

"We are used to admiring the achievements of science, the deep understanding . . . of the natural world that the scientific method of enquiry and reason has brought us", writes Scott, launching into his last chapter. Well, up to a point, Lord Copper, but I suspect that for most of Scott's intended public it comes fairly well down the agenda. Morowitz demands a whole lot less of his readers and may in the end make more friends — or at least nodding acquaintances — for science. This, his third volume, consists of *feuilletons*, reprinted from an organ called *Hospital Practice*, and addressed more (I suspect) to those who wield the stethoscope than the bedpan.

Morowitz is a biochemist and, it seems, a gourmet, traveller and *boulevardier*, and he comes over better in some personas than in others. There are times when the inspiration burns low: casting around for a topic, he finds it in his breakfast cup of coffee. He hastens to the library to turn up references to coffee but comes up with few surprises. I was moved at this point to seek out one of Robert Lynd's essays, written one Easter on the subject of eggs. When Lynd looked up 'egg' in a work of reference he found "Egg, Augustus Leopold (1816–63), English genre pain-

ter", and thus rewarded, he was away on a tide of fancy.

It is in his discourses on scientific subjects that Morowitz is at his best. His account of how studies on biological arcana, such as would have incurred the scorn of Senator William Proxmire, led directly to an understanding and treatment of a grievous disease, myasthenia gravis, could scarcely be bettered. His tale of the students who set up a business for supplying dried brine-shrimps to high-school laboratories for a range of diverting experiments and ran into commercial competition might have served H. G. Wells for a short story.

Morowitz's scepticism about medical fashions, in diet for instance, is bracing, and I liked his quotation from a small-town paediatrician of his acquaintance: "There aren't many children who need their tonsils out. There are otolaryngologists who need new cars or whose wives need new coats, but there aren't many children who need their tonsils out." Such procedures became known later, I believe, as remunerectomies. Morowitz is altogether too tolerant of blather from the likes of Teilhard de Chardin, and sometimes hands you a dog-biscuit like this — "Sexual dimorphism at the cerebral or intellectual level is so great that reality at the epistemological or ontological levels will be gender-dependent" — to make you reach for the bicarbonate. But his balance stands well in the black, and his chocolate-coated nuggets of science will continue to entertain and do surreptitious good. He is, moreover, a deft hand with the literary quote; but to toss in Alexander Pope's epitaph for Newton —

Nature, and Nature's laws, lay hid in night:
God said, *Let Newton be!* and all was light

— without J. C. Squire's rejoinder, is like dropping one shoe. Let me remind Dr Morowitz:

It did not last: the Devil, howling *Ho*,
Let Einstein be, restored the status quo. □

Walter Gratzer is in the MRC Muscle and Cell Motility, Kings College London, 26–29 Drury Lane, London WC2B 5RL, UK.

New in paperback

■ From Harvard University Press comes Linda Schiebinger's *The Mind has no Sex? Women in the Origins of Modern Science* (for review see *Nature* **342**, 309; 1989). Price is \$12.95, £10.25.

■ *Science à la Mode: Physical Fashions and Fictions* by Tony Rothman is a collection of essays about what happens when scientists jump on band-wagons (for review see *Nature* **340**, 608; 1990). Published by Princeton University Press, price is \$9.95.

■ In *Reliable Knowledge*, John Ziman explores the grounds for belief in science. Published by Cambridge University Press, price is £5.95, \$8.95. □

Fearless steps into darkness

J. H. Mulvey

Conversations on the Dark Secrets of Physics. By Edward Teller, Wendy Teller and Wilson Talley. Plenum: 1991. Pp. 247. \$23.95 (US), \$28.74 (rest of the world).

EDWARD Teller was a 19-year-old student at Karlsruhe in 1927 when Werner Heisenberg published the uncertainty principle. Soon afterwards, this bright Hungarian, who had often helped one of his lecturers over difficult points in the new quantum mechanics, left to



Teller — cause of some maths anxiety?

study under Heisenberg in Munich.

Now, at 83, he is known worldwide as the 'father' of the American hydrogen bomb, but that story and Teller's later role as the moving spirit behind Reagan's Strategic Defense Initiative are not the subjects of this book, although the notoriety they have earned him will spur many to consider buying it. Here, Teller gives his account of the steps leading to the revolution that established a new order in physics in the first quarter of this century. The origins are lectures in the appreciation of physics given at the University of Chicago over a 40-year period and, more recently, a course he gave to high-school students and their teachers.

He starts with the warning: "I will use mathematics because physics without mathematics is meaningless"; and, accepting that not everyone will be able to follow him in everything he says, remarks that he hopes to convey "a feeling" for what is going on.

The first chapter jumps straight into Einstein's special relativity theory by asserting the invariance of the quantity $(ct)^2 - r^2$ to transformations between sys-

tems in uniform relative motion. This leads to the constancy of c , the velocity of light, and an end to the classical universality of time intervals, t , for events separated by a distance r . The development is concise and it seems, as Teller says, that "relativity is simple". But he chooses not to follow the traditional approach of considering the experiences of two observers of the same events — one in a train and the other on the station platform — to prepare the reader for the unexpected failure of the 'common-sense' view that events simultaneous for one are simultaneous for all observers.

On the way to Newton via Aristotle, Copernicus, Brahe and Galileo, he picks out Kepler as a great man of physics, "a paragon of scientific honesty", for abandoning years of work on epicycles to try elliptic orbits for the planets. Although Newton claimed, "I make no hypotheses", his great synthesis of the motions of apples and planets begs one of the fundamental questions still not fully resolved in physics: why is the mass that appears in Newton's law for the force of gravitational attraction the same as the mass in his law for inertial forces: force = mass $\times$ acceleration? The principle of the equivalence of gravitational and inertial mass lies at the heart of Einstein's theory of gravitation, which is formulated in terms of curvature in a four-dimensional space-time; "do not be dismayed", encourages Teller on reaching this juncture, going on to say that he wants only to introduce the idea of curvature in four dimensions without the mathematical complications. After each chapter there are a few questions for the reader to try, with fully worked answers at the end of the book. This adds background to the discussion, but some demand considerable mathematical competence; one asks for an estimate of the deflection of starlight by the Sun.

Statistical mechanics introduces probabilities, the Boltzmann factor, the idea of increasing disorder and irreversible processes. Teller estimates how long it might take to sense the presence of a perfumed lady, and poses the question: "On a very hot day does it make sense to open the refrigerator door to cool down the house?"

Faraday discovered the links between electric current and magnetism, inventing the electric motor and generator. But also, of immense importance for the future development of physics, he introduced the fundamental concept of a 'field', a region of space populated by 'lines of force'. Now Teller, advising readers to "just get the general idea of what's happening", describes the differential operators 'divergence' and 'curl', and develops Maxwell's equations of electromagnetism. Throughout the

book, readers eavesdrop on "conversations" — rather one-sided — between Teller and his two assistants, daughter Wendy and Wilson Talley. These appear as footnoted exchanges, sometimes a little trivial but usually amplifying or extending the main text. In one, his daughter tells how Teller's father, a lawyer, had hoped his son would become a chemist, but Teller showed greater enthusiasm for physics. So father took son to see their relative in Vienna, the famous physicist Paul Ehrenfest, who asked the young Teller, "Do you know what a curl is?". "Yes sir", replied Teller. "Let the boy study physics", said Ehrenfest. On such events hang dramatic turnings in history.

When Democritus advanced the concept of the atom as the smallest indivisible part of a substance, Plato sent him to see the local doctor, Hippocrates. The atomic, and molecular, structure of matter was not fully accepted before about 1910 and soon after, following Rutherford's discovery of the nucleus and Bohr's model of the atom, came the revolution in mechanics without which the behaviour of matter on the atomic scale could not be understood. Teller emphasizes the importance of Bohr's 'correspondence principle' in the early development of quantum mechanics and uses it as a guide to understanding the form of the hydrogen-atom energy levels. This is not one of the clearest sections of the book: the argument is not well presented and one of the diagrams is not consistent with the text.

The penultimate chapter deals with the debates over the probability interpretation of quantum mechanics, arriving at the uncertainty relations and the still troubling problem, for many physicists, of the description — or rather lack of one — for the act of measurement: that instant when a wave-function allowing a particle to have probabilities of being anywhere in a large volume "collapses" to give the certainty of its observation at a particular spot.

As one of the few still living who worked and talked with Heisenberg, Bohr, Pauli, Dirac and other leaders of the revolution in those golden years of discovery, it is perhaps natural that Teller finds "no truly novel intellectual developments in the last sixty years". It is true that in those early years were laid down most of the foundations for the considerable advances in physics that have been made since, and the practical consequences — as examples he discusses semiconductors and lasers in the last chapter — are still a long way from being exhausted.

This is a challenging, original, very personal account of the fundamentals of physics. Apart from some lapses it is lucid, laced with humour, anecdotes and

insights into other physicists, particular Bohr. But many paragraphs are peppered with mathematical statements, and although Teller promises to explain his maths he sometimes forgets and the 'high-school student' will occasionally be faced with complex mathematical expressions without help. Reading such sections is not easy; unlike a member of an audience at a lecture, one is not swept past the tricky bits by the momentum of spoken words.

Teller has entered a domain where few have dared tread; he attempts a book that goes beyond recounting the major landmarks on the way to modern physics to show that these new concepts and understandings could not be reached without the use, and development, of the mathematical language in which they

are expressed. This is a bold assignment, but readers who are genuinely interested in what physics is about and who are willing to think as they read, rather than just look at coloured pictures — of which there are none — will gain a real appreciation of the power of physics, and an inkling of the way a physicist thinks.

But why "Dark Secrets"? Perhaps a clue is to be found in Teller's words about atoms: "We have seen the atoms. We cannot escape them. When you learn how they behave, you may wish that they did not exist." □

John Mulvey is in the Partical and Nuclear Physics Laboratory, Department of Physics, University of Oxford, Keble Road, Oxford OX1 3RH, UK.

Keeping in shape

Nick Cowan

The Neuronal Cytoskeleton. Edited by Robert O. Burgoyne. Wiley-Liss: 1991. Pp. 344. £62.80, \$94.95.

FROM time to time, I receive requests for a cytoskeletal probe (most often tubulin) to be used as an example of what is referred to as a 'housekeeping function'. This irritating epithet carries with it the implication that, in stark contrast to the exciting system under study, the components of the eukaryotic cytoskeleton are involved in biologically mundane tasks, and are therefore quintessentially dull.

Fortunately, such perceptions are based entirely on ignorance, not least because the cytoskeleton is involved in such diverse and vital functions as cell division, the acquisition and maintenance of cell shape, intracellular transport and motility. Moreover, with the suspicion that cytoskeletal abnormalities may have a role in the pathogenesis of some neurodegenerative disorders (for example, Alzheimer's disease), the cytoskeleton has now acquired a measure of respectability not often enjoyed by housekeepers. This collection of review articles dealing with aspects of the neuronal cytoskeleton is therefore both timely and appropriate.

Assembling a dozen or so chapters from different authors for a book covering topics varying in scope from ultrastructural morphology to gene expression is not an easy task. This is particularly so given the competition from review articles in weekly or monthly publications, not to mention the recent profusion of journals dedicated largely to distillations from the literature.

Although journal reviews are usually shorter and often less comprehensive, they do tend to be much more up to date. Perhaps this reflects no more than the slower turn-around time for a hard-back volume compared with a journal (one tardy contributor can gum up the whole works). Even so, the absence of some important new information in the book is conspicuous, despite some notes added in proof.

For example, there is no discussion of the evidence that neurofibrillary tangles contain tau protein or tau protein fragments — indeed, the index has only two entries for Alzheimer's disease, one of which refers to the introductory chapter emphasizing the promise that the study of the cytoskeleton holds for understanding neurological disease. The chapter on microtubule-associated proteins asserts that "no sequence data is available for any of the MAP1 species", although the complete sequence of MAP1B and a study of the microtubule-binding properties of its repeated motifs were published in 1989. And in a section on post-translational modifications of tubulin, there is no mention of polyglutamylation; this apparently unique process was described in January 1990, and accounts for the array of beta-tubulin spots seen on two-dimensional gels of whole extracts of brain.

Despite such shortcomings, there is a wealth of useful information in this well-produced book, which should live up to its claim of being valuable to researchers, clinicians, postdoctoral fellows and graduate students in neurobiology, if only in the short-term. Those interested in housekeeping functions, on the other hand, should perhaps turn their attentions elsewhere. □

Nick Cowan is in the Department of Biochemistry, New York University Medical Center, 550 First Avenue, New York, New York 10016, USA.

The spread of HIV-1 in Africa: sexual contact patterns and the predicted demographic impact of AIDS

R. M. Anderson, R. M. May, M. C. Boily, G. P. Garnett & J. T. Rowley

The spread of HIV-1 in Africa is examined here in the light of recent information on the main epidemiological and behavioural determinants of transmission. Mathematical models incorporating demographic, epidemiological and behavioural processes are used to assess the potential demographic impact of the disease AIDS. These analyses highlight the significance of patterns of sexual behaviour, and in particular networks of sexual contact, on the predicted spread of infection. Current data reveal substantial variations in the degree of spread between and in countries, but new analyses support earlier predictions that in the worst-afflicted areas AIDS is likely to change population growth rates from positive to negative values in a few decades.

TEN years after the first description of AIDS¹, the spread of its aetiological agents (the human immunodeficiency viruses HIV-1 and HIV-2^{2,3}) continues unabated. In the worst-afflicted regions, such as sub-Saharan Africa, the pattern of spread of these lethal viruses reveals a depressing picture with steadily increasing levels of HIV-1 and HIV-2 infection in the general heterosexual population and very high amounts of infection in the major at-risk groups such as female prostitutes, their male clients and patients attending sexually transmitted disease clinics. In certain urban centres, AIDS is now the leading cause of mortality in adults⁴ and one of the main determinants of infant mortality⁵. The extent of spread of HIV-1 in Africa, based on numerous serological surveys over the period 1985–90, is summarized in Fig. 1 and Table 1a. In the rest of the world the pattern of spread is very variable between regions, with encouraging evidence of a decreasing rate of growth of the epidemic in male homosexuals in developed countries (after an initial period of rapid spread from 1981 to 1986)⁶, continuing rapid spread among intravenous drug users in Europe, North America and parts of South East Asia^{7,8}, and a slow but steady rise in heterosexual populations in certain developed countries, South America and parts of Asia⁹.

Despite much research, there remain many uncertainties in the chief epidemiological and behavioural parameters that influence transmission, due in part to the long and variable incubation period of the disease¹⁰, to the genetic variability of the virus¹¹ and to the difficulties that surround the study and quantification of human sexual behaviour¹². Here we examine some of these uncertainties, emphasizing the role of sexual behaviour and contact patterns in determining viral spread, the potential demographic impact of AIDS in Africa, and progress in the development of mathematical models that combine demographic and epidemiological processes in pursuit of better interpretation and prediction.

Epidemiology in Africa

The principal features of HIV-1 spread in Africa are summarized in Fig. 2. Most horizontal transmission events between sexually active adults are through heterosexual contact, with vertical transmission from infected mother to infant being most significant in initiating infection in the youngest age groups. In the general adult population, infection has spread steadily since 1984, reaching 20–30% in the worst-afflicted urban centres as reflected by serological surveys amongst pregnant women and blood donors (Fig. 2a). The distribution of infection by age and sex rises steadily as teenagers begin sexual activity, with

maximum HIV-1 prevalence in 20–25-year-old women and 25–35-year-old men (Table 1b; Fig. 2b). This difference is thought to reflect sexual contact of older men with younger women. In the case of HIV-2, peak prevalences are observed in older age classes but the maximum still occurs in women at earlier ages than in men (Fig. 2b). The more even spread of HIV-2 infection in older age classes may result from a longer incubation period before disease and subsequent mortality when compared with HIV-1, and/or may reflect a long duration of viral spread in the Guinea-Bissau population referred to in Fig. 2b.

The rate of spread in the general urban population differs greatly between regions, being rapid in Malawi, Rwanda, Uganda, Tanzania and Zambia and much slower in Kenya, Mali and Zaire (Table 1b). In high-risk groups, such as female prostitutes in Nairobi, transmission has been very rapid (Fig. 2d), but not as rapid as in certain intravenous drug-using communities in North America or Thailand (Fig. 2e). Patchiness in transmission is reflected by differences in urban and rural areas, such as in Tanzania (Fig. 2c), with higher amounts of infection in the former by comparison with the latter. During the past 2 years, however, there has been a steady rise in prevalence in many rural regions (Fig. 1b). The differing amounts of infection depend on various factors including the time since the introduction of the infection in a defined area, rates of sexual partner change, patterns of sexual contact between age classes and between individuals in urban and rural populations, and the frequency with which males have sexual contact with female prostitutes. Estimates of the doubling time of the epidemic, as reflected by longitudinal changes in HIV-1 seroprevalence, range from 1 to 3 years in the general populations in the worst afflicted regions, to 5 years or longer in areas where the prevalence is currently low. In female prostitutes in urban centres the doubling time is often 1 year or less (Table 2a). But the general pattern in Africa is unclear at present because of poor disease-notification systems (and the fact that at present AIDS cases reflect the rate of viral spread in the highest-risk groups) and the scarcity of reliable longitudinal data (ideally based on cohort studies) on seroprevalence in defined groups (the four most detailed data sets are shown in Fig. 2a).

Epidemiological parameters

Information on the main epidemiological parameters determining transmission is slowly accumulating but much uncertainty remains. The major features of infection and disease culled from some of the more extensive studies are summarized in Table 2. In male homosexuals and adult transfusion recipients, roughly

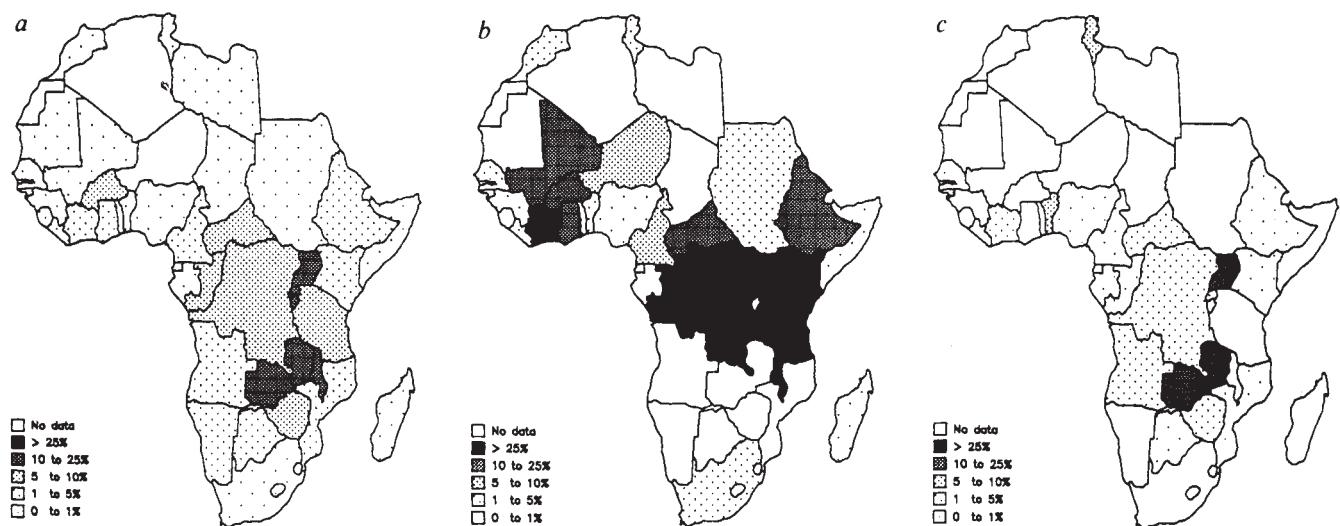


FIG. 1 Maps recording estimates of the prevalence of HIV-1 infection in the heterosexual populations of different African countries. *a*, Sexually active population in urban settings; *b*, female prostitutes (largely in urban settings); *c*, sexually active population in rural settings. The maps are based on numerous published and unpublished surveys of seroprevalence in the general population, pregnant women, blood donors, various occupational groups and female prostitutes recorded in the HIV/AIDS surveillance database⁴⁹. They cover the period 1985–90 and hence probably underestimate the degree of infection in early 1991 as many of the surveys used

to compile the maps are 2 or 3 years out of date. The patterns must be viewed with caution given the wide variability in study design and quality between the many surveys used to estimate average prevalence in the defined risk groups. The figures given for a particular country are the weighted (by sample size) average value of all surveys recorded in the database, as detailed in Table 1a. The general population refers to seroprevalence in surveys in non-high-risk groups such as pregnant women, blood donors, certain occupational groups and the community in general.

50% have developed AIDS by year 10 of infection, leading to estimates of the mean incubation period (time between first infection to the development of the disease) of between 8–10 years, on the assumption that most of those infected will develop AIDS (which appears likely). This period appears independent of sex or risk-group, but dependent on age¹⁰ and, perhaps, geographical region. There is some evidence from Africa to suggest a much shorter mean period in female prostitutes¹³, but other studies do not reveal significant differences from the rates of progression recorded in developed countries¹⁴. In perinatally infected infants the mean period seems to be between 1 and 2 years¹⁵. Better treatment and management of asymptomatics in developed countries seems to be increasing the average incubation period¹⁶. Evidence is emerging that the highly variable incubation period may in part be due to unpredictable (mutation) events determining the emergence of new viral quasispecies with variable cytopathic properties, cell tropisms and replication rates in an infected person¹⁷, and that serious immunodeficiency may be due to viral quasispecies diversity exceeding some threshold beyond which the human immune system is unable to control viral replication¹⁸.

Survival of infants once AIDS (or HIV-related serious disease) is diagnosed is typically less than 1 year, whereas it is between 1 and 2 years on average, and increasing, in adults in developed countries (Table 2f). HIV-1-infected mothers in Africa are more likely to have infants that die at or before birth than are uninfected women¹⁹. Of live births to infected mothers, roughly 40% of infants seem to acquire infection, on the basis of studies in Africa, or roughly 20–30% on the basis of studies in developed countries (Table 2e). Much uncertainty surrounds these estimates as some infants remain antibody-positive and antigen-negative, others change from antibody-positive to antibody-negative, whereas others are antibody-negative and antigen-positive^{19–23}.

Data on the rate of horizontal transmission through sexual contact between heterosexuals is very limited at present because of difficulties in study design and the need to quantify the various facets of sexual behaviour that influence transmission (for

example, type and frequency of sexual contact). The transmission probability may be defined per sex act or per partnership, with most studies focusing on the latter with, unfortunately, little information on the likelihood per unit of time, per act or per partner contact. The average pattern recorded in most published studies (see Table 2d) suggests transmission probabilities per partnership of 20% from male to female (β_1) and 11% from female to male (β_2). Great variability is apparent in transmission studies and some surveys record equal probabilities²⁴. But the apparent average twofold difference between the probabilities of male-to-female and female-to-male transmission may in part explain the observed bias in the sex ratio of HIV-1 infection in Africa where more women seem to be infected than men (1:1.4 male to female ratio)^{25,26}. Simple theory suggests that if the effective mean rate of sexual partner change for men is equal to that for women, then the male-to-female ratio of HIV infection is roughly given by $(\beta_2/\beta_1)^{1/2}$ (ref. 27). With values of $\beta_1 = 0.11$, $\beta_2 = 0.2$ (Table 2) this gives a ratio of 1:1.35, which is close to that observed. The likelihood of vertical and horizontal transmission is positively correlated with the severity of disease symptoms in the mother or infected sexual partner (in part reflected by viral or antigen concentration)²⁸. It may also be linked with the sporadic emergence of new quasispecies of the virus in the infected person, with the appropriate biological properties that permit transmission.

Sexual behaviour

The higher prevalence of HIV-1 in heterosexual populations in Africa (by comparison with developed countries) has been linked to its having been spreading for longer there, to much higher prevalences of untreated sexually transmitted diseases (STDs) such as genital ulcers that may promote HIV transmission²⁹, and to higher rates of sexual partner change in African societies^{30,31,90}. Evidence for the first two explanations is growing but that for the third is limited at present (Table 2g). The World Health Organization has initiated widescale surveys (KABP questionnaire surveys³²) on sexual behaviour in Africa and data are beginning to accumulate with surveys in 10 coun-

TABLE 1 Prevalence of HIV in Africa

(a) Weighted mean percentages of the 'general' sexually active urban population with antibodies to HIV-1, based on ELISA tests and other summary statistics

Country	Weighted mean per cent seroprevalence*	Number of surveys	Total number of people tested	Range of seroprevalence estimate	Year of last survey	Estimated urban population size (x100,000)
Algeria	No data					102
Angola	3.6	22	2,366	0.0-16.67	1988	23
Benin	0.12	7	5,046	0.0-0.28	1987	17
Botswana	3.0	3	405	0.0-4.28	1987	2
Burkina Faso	7.3	6	2,222	0.0-10.0	1989	7
Burundi	15.2	2	1,018	4.3-16.3	1986	4
Cameroon	1.0	58	65,523	0.0-10.8	1989	50
Cape Verde	0.04	7	2,496	0.0-0.11	1988	—
C.A.R.	7.5	40	18,177	0.0-17.07	1989	12
Chad	0.0	2	707	—	1986	16
Comoro Islands	0.0	3	995	—	1988	—
Congo	5.7	26	38,864	0.0-24.2	1989	8
Cote d'Ivoire	4.7	68	31,268	0.0-12.99	1989	49
Djibouti	0.2	5	3,336	0.0-2.44	1988	—
Egypt	0.1	16	47,134	0.0-3.7	1989	240
Equatorial Guinea	0.3	2	722	0.26-0.3	1988	—
Ethiopia	2.0	12	14,552	0.0-8.17	1990	54
Gabon	0.5	10	6,559	0.0-1.83	1988	5
Gambia	0.1	3	9,806	0.0-0.12	1989	—
Ghana	3.3	4	1,086	0.0-4.63	1987	44
Guinea	0.3	22	11,049	0.0-1.68	1988-89	16
Guinea-Bissau	0.4	19	11,813	0.0-7.4	1988	—
Kenya	3.4	23	95,335	1.5-25.33	1990	49
Lesotho	0.09	6	10,465	0.0-2.2	1989	3
Liberia	0.3	7	3,929	0.0-0.55	1989	10
Libya	0.0	2	3,064	—	1986-87	27
Madagascar	0.03	10	16,610	0.0-0.04	1989	25
Malawi	17.0	19	7,736	0.0-23.26	1989	10
Mali	0.2	6	1,858	0.0-0.35	1988	15
Mauritania	0.1	13	1,618	0.0-0.41	1987-88	7
Morocco	0.02	5	5,774	0.0-0.03	1984-87	11
Mozambique	3.4	53	44,183	0.0-12.61	1989	34
Namibia	2.1	2	851	0.0-2.54	1988	5
Niger	No data					12
Nigeria	0.3	43	290,302	0.0-1.48	1989	352
Rwanda	21.4	30	11,273	2.98-31.86	1990	4
Sao Tome & Principe	1.0	2	200	0.0-2.0	1988	—
Senegal	0.1	32	27,884	0.0-1.5	1989	26
Sierra Leone	3.5	1	285	—	1987-89	10
Somalia	0.02	9	3,390	0.0-0.07	1988	21
South Africa	0.04	95	5,705,339	0.0-3.76	1990	189
Sudan	0.3	9	31,593	0.0-2.91	1989	49
Tanzania	8.5	47	16,028	0.0-32.81	1989	69
Togo	No data					9
Tunisia	0.08	4	2,667	0.0-0.14	1985-87	41
Uganda	15.2	40	46,208	0.0-33.3	1989	16
Western Sahara	No data					—
Zaire	5.4	83	134,910	0.9-13.42	1989	124
Zambia	13.2	43	19,835	1.96-37.5	1989	38
Zimbabwe	5.6	6	2,110	0.0-36.7	1989	23

(b) Seroprevalence of HIV-1 or HIV-2 in blood donors, pregnant women or the general population in urban areas in Africa, 1987-89

Country	City	Year	Risk group	Sample size	Prevalence (%)	Reference
(HIV-1 or 2)						
Guinea-Bissau	Bissau	1989	Pregnant women	2,000	(2) 7.0	50
Cote D'Ivoire	Abidjan	1989	Blood donors	2,083	(1) 7.4	51
Kenya	Nairobi	1989-90	Pregnant women	726	(1) 5.4	52
Malawi	Blantyre	1989	Pregnant women	126	(1) 18.3	53
Mali	National survey	1988	Blood donors	1,841	(1) 4.1	54
Mozambique	Maputo city	1989	Blood donors	1,380	(1) 12.6	55
Rwanda	Kigali	1989	Pregnant women	900	(1) 30.3	56
Tanzania	Bukoba	1987	General population	573	(1) 32.8	57
Uganda	Kampala	1989	Pregnant women	650	(1) 31.7	58
Zaire	Kinshasa	1989	Pregnant women	8,108	(1) 6.0	5
Zambia	Lusaka	1989	Blood donors	5,505	(1) 18.8	59

* Weighting based on the sample size of each survey⁴⁹.

tries recently completed. The main features of the data are great variability in reported behaviours between and in countries, somewhat higher rates of sexual partner change than those reported in developed countries (perhaps by a factor of 2; see Table 2), higher rates reported by men in comparison with women (perhaps due to exaggeration by men, or to the failure to take account of the sexual contacts made by men with female

prostitutes), marked changes in sexual activity with age, and men on average forming sexual partnerships with women younger than themselves^{33,34}. The last factor is of particular importance in assessing the potential demographic impact of AIDS, as it acts to intensify transmission in young sexually mature females who potentially contribute most to the net fertility of the population³⁵. A recent study in Guinea-Bissau, for

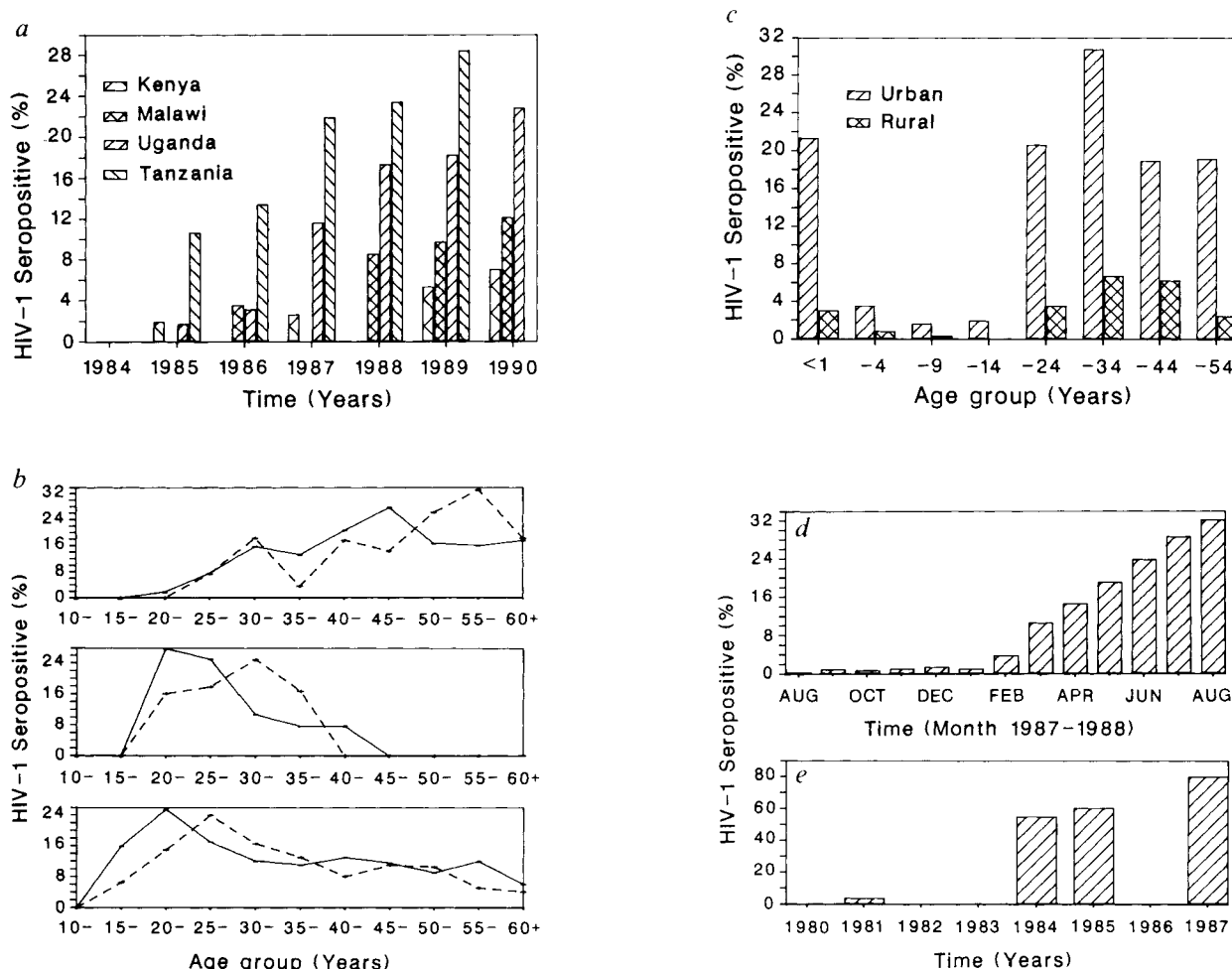


FIG. 2 Epidemiological patterns of HIV-1 and HIV-2 infection in Africa. *a*, Longitudinal (noncohort) trends in HIV-1 seroprevalence over the period 1984-90 in urban pregnant women in Kenya, Malawi, Uganda and Tanzania⁴⁹. *b*, Age and sex stratified seroprevalence surveys of HIV-1 in the general population in Uganda in 1988 (bottom)⁸⁶, Zambia in 1985 (middle)⁸⁷ and HIV-2 in Guinea-Bissau in 1987 (top)⁸⁸. Male, ---; female, —. *c*, Age

prevalence of HIV-1 in urban and rural populations in 1987 in the Kagera region of Tanzania⁵⁷. *d*, Spread of HIV-1 in intravenous drug users attending Thanyorak Hospital, Bangkok from August 1987 to August 1988⁸⁹. *e*, Spread of HIV-1 in female prostitutes in Nairobi from 1980 to 1987⁶⁰. Spread is rapid but not as rapid as that for intravenous drug users in Bangkok (*d*).

example, recorded a mean difference of age between man and spouse of 9 years³³. Heterogeneity in rates of sexual partner change in African societies appears similar in magnitude to that recorded in developed countries where average behaviour (m) is related to variability in behaviour (σ^2) by a power law, $\sigma^2 = am^b$, where a and b are coefficients and the value of b is about 3 (ref. 36). The uniformity in this pattern across widely different societies is difficult to explain. It may be linked to a set of basic rules governing partner formation and dissolution, common to all societies, but the pattern simply states that average behaviour is positively (and supra-linearly) associated with the extremes of behaviour in a given community.

As well as partner change rates and frequency and type of sexual contact in partnerships, the most important processes that determine the pattern of HIV transmission are those governing mixing in or between different strata of the population³⁷. These strata include age classes, sexual-activity classes and spatial location (for example urban, rural). There is no information on these processes apart from the general trend for men to form sexual partnerships with younger women. The issue of matching supply and demand for partnerships between women and men is linked to the problem of determining networks of sexual contacts or mixing patterns. For example, if AIDS-induced mortality reduces the number of highly sexually active women (for example, female prostitutes), do men reduce their

rates of sexual partner contact, do low-activity women increase activity to meet the unfilled demand, or do new female recruits to the prostitute population offset the losses induced by AIDS? In the worst-affected populations, where AIDS is a chief cause of mortality, these issues remain unresolved despite their obvious importance to the future pattern of the epidemic.

Mathematical models

Early work on mathematical models that incorporate epidemiological and demographic processes suggested that AIDS is capable of turning positive population growth rates, of the magnitude of 3% per annum or more, negative over timescales of a few to many decades²⁷. The disease selectively acts to increase mortality in young infants (through vertical transmission) and in the sexually active and productive age groups. Models suggest that the overall ratio of productive adults to infants and children (the dependency ratio^{27,31,35}) alters little under the impact of AIDS, but at the family level parentless or childless groupings generated by variation in the time sequence of infection and subsequent mortality will create serious social problems. Much controversy surrounded these predictions³⁸, mainly because many of the early models did not adequately account for heterogeneity in sexual contact between different strata in the population. In the past 2 years, models have been developed that address various aspects of the observed heterogeneity in

TABLE 2 Summary of selected information on the major epidemiological and behavioural parameters that determine the rate of transmission of HIV and the demographic impact of AIDS

(a) Rate of spread of infection as measured by the doubling time (yr) for the prevalence of HIV-1 infection in defined populations in African countries					(e) Likelihood of vertical transmission (percentage of babies born to infected mothers with evidence of HIV-1 infection)					
Country	Risk group	Time period	Doubling time (yr)	Ref.	Country	Sample size	% with HIV infection or antibodies	Ref.		
Kenya	Pregnant women	1985-89	2.8	49	USA	59	17	77		
Uganda	Pregnant women	1985-89	2.8	49	Europe	271	24	66		
Malawi	Pregnant women	1985-89	1.2	49	USA	55	29	78		
Kenya, Nairobi	Prostitutes	1981-85	<1.0	60	Italy	486	33	20		
(b) Sex ratio (male to female) of cases of HIV-1 infection					USA	117	35	21		
Uganda	General population	1987	1:1.4	25	Zaire	92	39	5		
Zaire, Kananga	Hospital patients	1988	1:1.5	26	Zambia	109	39	19		
(c) Rates of progression from infection to the development of AIDS (a measure of the incubation period of AIDS, and the duration of infectiousness) based on cohort studies					Europe	372	13-15	22		
					(f) Survival once AIDS is diagnosed					
					ADULTS					
Country	Risk group	Years since infection	Per cent with AIDS	Ref.	Country	Sample size	Years	Median survival in months	Ref.	
ADULTS					UK*	997	1981-88	10.36	79	
USA	Male homosexuals	11	54	61	USA*	5,833	1981-87	11.4	80	
USA	Transfusion recipients	8	49	62	HIV-1 INFECTED INFANTS BORN TO INFECTED MOTHERS					
USA, Europe, Australia	Male homosexuals	7	39	10	Country	Sample size	Mortality		Ref.	
USA, Europe, Australia	Adult haemophiliacs	7	27	10	Zambia	42	44% at year 2		19	
USA, Europe, Australia	Child haemophiliacs	7	20	10	USA	172	Median survival 3.2 years		23	
Sweden	Transfusion recipients	5	22	63	(g) Sexual behaviour recorded as the claimed number of different sexual partners per defined period of time (heterosexuals, general population surveys)					
Italy	Intravenous drug users	4	18	64		Sex	Sample size	Mean (yr ⁻¹)	Comment	Ref.
Kenya, Nairobi	Female prostitutes	3.3	50	13	UK	Both	823	1.5		81
INFANTS					UK	M	337	1.1		82
USA	Perinatally infected	1.5	59	15	UK	F	443	0.89		82
USA	Perinatally infected	1.0	20	65	USA	Both	851	2.1		83
Europe	Perinatally infected	1.25	42	66	Uganda	M	42	2.8	20-34-yr olds	84
(d) Likelihood of horizontal transmission (percentage who acquire HIV-1 via sexual contact with infected spouse)					Uganda	F	51	1.5	20-34-yr olds	84
MALE TO FEMALE					Kenya	M	132	11.8		30
					Rwanda	M	27	3.0		85
					Cote d'Ivoire	Both	—	2.3	Single persons	32
					Uganda	M	214	4.0	Rural	86
					Uganda	F	505	1.0	Rural	86
					Guinea-Bissau	M	152	2.5		33
					Kenya	M	194	9.3		30
					Kenya	F	112	1.2		30

* Survival times are increasing in developing countries due to the use of AZT and better management of opportunistic infections.

sexual behaviour^{31,35}. But theoretical developments have out-paced data availability, given the many practical difficulties associated with the study of sexual behaviour and, in particular, networks of sexual contact. Even so, the theory provides important clues to the interpretation of observed patterns, suggests what data must be acquired, and provides a framework to assess the impact of different control interventions.

Heterogeneity in virus transmission may arise from spatial factors (contacts between villages or urban and rural centres); nonhomogeneous mixing between the age classes of the two sexes or between different social classes, age-dependent changes

in sexual activity; and different patterns of mixing between the different sexual activity classes (defined on the basis of rates of sexual partner change). The last factor is of particular importance in determining HIV-1 spread in male homosexual communities in developed countries, where like-with-like (assortative) mixing generates rapid spread, but with infection largely restricted to the small population of highly sexually active individuals, whereas like-with-unlike (disassortative) mixing generates slower spread but a more widely disseminated epidemic^{39,40}.

In the context of the spatial diffusion of infection, simple

models of sexual contacts in and between villages in rural locations have been used to examine the period over which HIV-1 has been spreading in Africa. Analyses of molecular sequence data from lentiviruses of monkeys (simian immunodeficiency virus, SIV) and humans (HIV-1 and HIV-2) suggest that the human viruses diverged from the monkey strain at least 140–160 years ago^{41,42}. Conventional, homogenous mixing epidemiological models find this fact hard to explain (unless R_0 lay infinitesimally above unity for most of this century-long period), but the more complex in-and-between village transmission models suggest that from the initial stuttering chains of transmission events in one village (in which the basic reproductive rate of infection, R_{01} , could be less than unity²⁷), the virus can persist through a network of limited sexual contacts between villages, until the combined rate of between- and in-village transmission is sufficient to generate a major epidemic which will only become apparent after a century or more, once infection levels have risen to measurable levels. An illustration of such a calculation is presented in Fig. 3a, where for contacts in any one 'seeding' village, $R_{01} < 1$, but overall $R_0 > 1$, due to the summed contributions of contacts between and in villages⁴³. In this example, HIV-1 seroprevalence takes 120 years to rise to 20% in the rural villages.

Once the virus has become established, after these initial stuttering chains of transmission events, the doubling time of the epidemic in the general population in the worst-afflicted regions (based on HIV-1 infection, not AIDS cases), appears to lie between 1 and 3 years (see Table 2a). Its value will be

determined by many factors, but its magnitude will greatly influence the time taken for AIDS-induced mortality to reverse the sign of population growth rates (Fig. 3b). Using the model in refs 27, 31, 35, a 1.5-year doubling time reversed the sign of a population growth rate that was 4% before the introduction of HIV-1 around year 30, whereas with a 2.5-year doubling time the period was around 60 years. One of the most important factors determining the magnitude of the net force of infection in the general population is the pattern of sexual contact between the different age classes of the two sexes. Models that take account of the observed bias for men, on average, to form sexual partnerships with women at least 5 years younger than themselves, suggest greatly enhanced demographic impact, when compared with predictions based on restricted in-age-class contact, all other factors being equal³¹ (Fig. 3c). This results from the concentration of infection in young women of childbearing age and the concomitant impact on infant mortality from AIDS through vertical transmission. This factor, combined with a greater likelihood of transmission from men to women than from women to men (see Table 2) generates predicted age and sex distributions of AIDS cases similar to those observed (Fig. 2b and 3d). The only major discrepancy lies in a greater number of predicted AIDS cases in young children than that observed. This is probably a consequence of the difficulties surrounding the diagnosis of HIV-1-related infant deaths in regions where other infectious diseases are a major influence on child survival.

The precise patterns of AIDS or HIV-1 infection by age and sex depend critically on the assumptions made concerning the

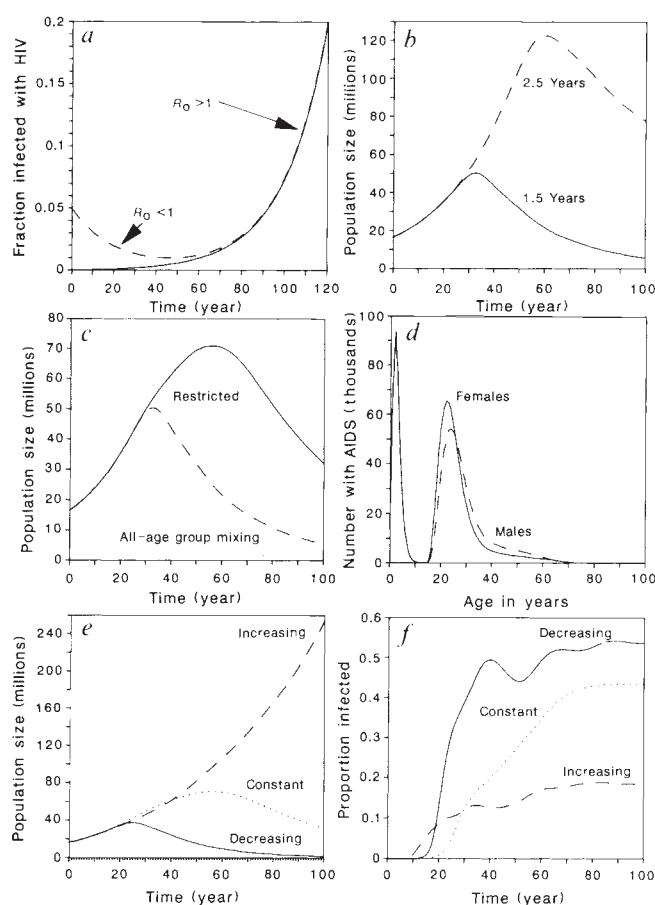


FIG. 3 Mathematical models and predicted patterns of the spread of HIV-1 plus the demographic impact of AIDS. *a*, Predicted temporal changes in the fraction infected with HIV-1 in the village into which the virus was first introduced (the 'seeding' village) and surrounding villages (average village) where sexual contact is predominantly in village but with limited between-village contact⁴³. The dashed curve represents the prevalence in the 'seeding' village whereas the solid line denotes the average value in surrounding villages. The basic reproductive rate R_0 was below unity for in-village transmission alone, although above unity for both in and between-village transmission. The initial decline in the levels of infection in any one seeded village, against a rising tide of diffusion of infection among villages is described in ref. 43. *b–f*, Simulations of the demographic impact of AIDS and the spread of HIV-1 infection in a population of 16.6 million (1:1 sex ratio) with a 4% per annum growth rate at time $t=0$ when HIV-1 was introduced (at a prevalence of 0.1%). The mathematical model, described in refs 27, 35 and 31, is age and sex structured and assumes that the average incubation period of AIDS in adults is 8 years (with variable infectiousness throughout this period) and 2 years in infants infected perinatally, and that the rate of vertical transmission is 50% (a high figure to reflect a high frequency of still births in infected mothers). Male to female transmission was assumed to be a factor of 3 greater than female to male and the probabilities of transmission were adjusted to mirror different doubling times of the epidemic in its early stages in the general population (1.5 years in *c–f*, 1.5 years and 2.5 years in *b*). Rates of sexual partner change per annum were varied in the different simulations as described for each graph as were patterns of sexual contact between the different age classes of the two sexes. Age-specific mortality and fertility rates were set at the start of the simulations to mirror the patterns typically observed in African countries and to give a net population growth rate of 4% in the absence of infection. *b*, Predicted demographic impact on total population size over a period of 100 years for forces of infection on the basis of 1.5-year or 2.5-year doubling time in the early stages of the epidemic. Mixing occurred between all age classes but the pattern was set to mirror men, on average, forming sexual partnerships with women 5 years younger than themselves. The rate of sexual partner change was independent of age (3.4 per year at the start of the epidemic). *c*, Influence of the pattern of mixing between the age classes of the two sexes on total population size. In the simulation labelled restricted, males and females only have sexual contact with partners in their own age class. In the simulation labelled all-age group mixing, males on average form partnerships with women 5 years younger than themselves. All other parameters were identical in the two simulations (mean rate of partner change independent of age at 3.4 yr⁻¹). *d*, Age and sex distribution of people with AIDS generated at year 50 in the all-age group mixing simulation in *c*. The pattern generated is very dependent on the assumptions made concerning age-dependency in rates of sexual partner change and the pattern of mixing between the age classes. *e*, Influence of age-dependent changes in the rate of sexual partner change (the same for women and men) on population growth under the assumption of restricted mixing. Three different assumptions are recorded, the mean increasing with age, the mean constant across age classes and the mean decreasing with age. In all cases the mean over all age classes was held constant at 3.4 yr⁻¹ at the start of the three simulations. Aside from age-dependency in sexual activity all other parameters were identical in the three projections. *f*, Predicted temporal trends in the proportion infected with HIV-1 for the three simulations in *e*. The detailed assumptions made in these numerical projections are described in refs 27, 31 and 35.

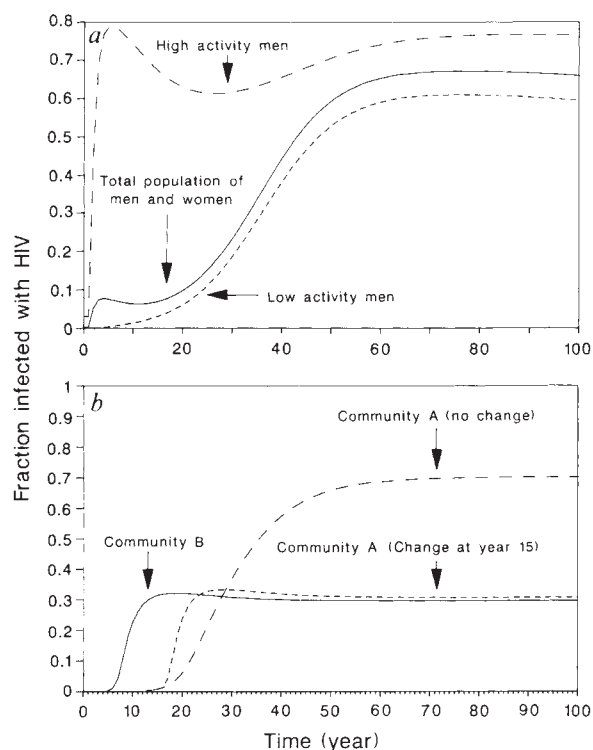
incubation period of the disease, sexual contact between age classes and age-related changes in sexual activity. Data on the last factor suggest that rates of sexual-partner change are highest in young sexually mature adults and decline in later adult life (perhaps after marriage), with the decline being slower in men than women^{33,34}. The precise pattern of age-related change has a large influence on the predicted demographic impact (Fig. 3e). In these predictions (in a population of size 16.6 million and a 4% per annum growth rate, at time $t=0$ when HIV-1 was introduced) the mean rate of sexual-partner change over all age classes was held constant at 3.4 partners per year, but in the three different simulations the age-specific mean was assumed to be constant, to increase with age, or to decrease with age. The third situation (that observed in survey studies^{33,34}) resulted in the greatest demographic impact, as a consequence of focusing infection in young women of childbearing age³¹ and in young sexually mature men. The associated temporal changes in the prevalence of infection are recorded in Fig. 3f. Note that in the simulation where the mean is highest in young adults, it is predicted to take 30 years or more before the level of infection approaches 50%, after the establishment of HIV-1 in the population.

A further important source of heterogeneity arises from the pattern of mixing between the different sexual activity classes (based on rates of sexual partner change). The simplest of models, based on the stratification of the population by sex and on two classes of sexual activity (low and high rates of sexual-partner change), illustrates this point⁸¹. For example, if mixing is set to mimic high-activity men having greatest contact with high-activity women (perhaps female prostitutes), but some contact with low-activity women (who have most contact with low-activity men), then a multiply peaked epidemic may occur,

with a rapid epidemic in the small proportion of high-activity men and women and a slower, but larger, longer-term epidemic in the low-activity men and women who constitute the majority of the population (Fig. 4a). The epidemic in the high-activity classes serves to seed the slower growing, longer-term epidemic. Similar patterns may arise with only one activity class of men with moderate levels of sexual activity (but higher than the low-activity class women), where female prostitutes serve to seed the epidemic in most monogamous women through contact with promiscuous husbands. The situations mirrored by these simple models may reflect what is occurring in cities such as Nairobi, where at present levels of infection are high in female prostitutes (60–80%), moderate to high in their male clients (20–40%), and low in pregnant women (5–6%). The low levels in the last group may continue to rise in the next decade, reflecting a second and much larger epidemic in the general population.

The subtle interplay between rates of sexual-partner change, patterns of mixing between sexual activity classes, and the balance between supply and demand of sexual contacts between the sexes could in some circumstances, trigger perverse outcomes if education efforts reduce rates of sexual-partner change in the general population. Consider, for example, the situation where one community has liberal attitudes to the formation of sexual liaisons (such that most people of both sexes have several sexual partners each year and, as a direct result, men have very limited contact with female prostitutes) and a second community in which most women are monogamous but men have many sexual partners per year through frequent contact with prostitutes⁸¹. If education in the former community reduces partner change rates among women, but, concomitantly, induces men to have more frequent contact with prostitutes, then the result may be an

FIG. 4 The influence of sexual contact patterns between different sexual activity classes of men and women (defined on the basis of the rate of sexual partner change per annum) on the pattern of the epidemic. The projections are based on a two sex, two sexual activity class (high and low rates of partner change) model described in ref. 81. *a*, Results of a simulation in which high-activity men (20% of the male population) have frequent sexual contact with high-activity women (prostitutes who form 1% of the female population) and limited contact with low-activity women (99% of the population) (assortative mixing with 99.5% in activity-class contact and 0.5% between activity-class contact). The mean rates of partner change for low- and high-activity men and women were set at 5.0, 20.0 and 4.0, 400.0 per annum, respectively (overall mean for men and women of 8 yr⁻¹). The probabilities of transmission per partner were set at 0.05 for women to men and 0.10 for men to women (yr⁻¹) and the vertical transmission efficiency was set at 40% (incubation period of AIDS of 8 years, with one year survival of AIDS patients, population size of 830,000, a 1:1 sex ratio, and the simulation was started with 4% of high-activity women infected and 3% of high-activity men). The three trajectories record temporal changes in the proportion infected with HIV for the total population (men and women) and high- and low-activity men. *b*, temporal changes in HIV infection in two communities. In community A, the majority of women have several different sexual partners per year such that male contact with very high-activity women (prostitutes, 3.6% of the population) is limited. The mean rates of partner change for low- and high-activity men and women were set at 7.0, 7.0 and 5.0, 60.0 per annum, respectively. (Other parameters are as defined in *a* excepting population size was set at 2×10^6 , male to female and female to male transmission probabilities were set equal to 0.05 yr⁻¹ and the simulation was started by the introduction of 10 infected high-activity women.) Two simulations are recorded one of no change in behaviour, and one where at year 15, low-activity women reduce their rate of partner change from 5 to 1 yr⁻¹, men keep their rate at 7 yr⁻¹, whereas high-activity women increase their rate to 180 yr⁻¹ to meet the increased demand from men created by the change in behaviour in low-activity women. In community B most women (96.4%) have few sexual partners (low activity is 1 partner yr⁻¹) while a few (3.6%) have very many partners (prostitutes with an average of 165.7 partners yr⁻¹). All men have moderate sexual activity (7 partners yr⁻¹) (all other parameters as defined for community A). Note that the spread of infection is limited (rising to a plateau of 30%) and restricted largely to high-activity women and some men.



acceleration in the rate of spread of the virus in the general population in the shorter term. In the longer term, however, the benefits will become apparent by a reduced overall size of the epidemic (Fig. 4b). In this example, unless one is aware of the influence of education not only on mean levels of sexual activity but also on patterns of mixing between risk groups, in the shorter term it could be falsely assumed that education was having a detrimental influence on the spread of infection! The uncertain effects of such changes in behaviour on patterns of mixing is somewhat analogous to the problem of understanding how sexual behaviour and mixing are likely to alter as a consequence of AIDS-induced mortality, which is severe in the high-sexual activity classes (that is, matching supply with demand).

Discussion

The current generation of mathematical models of the spread of HIV-1 in Africa reveals a bewildering array of possible outcomes depending on the precise details of sexual behaviour and patterns of sexual contact prevailing in a defined population. The assumed pattern of sexual behaviour dominates the predicted outcome more than small refinements associated with the rates of vertical or horizontal transmission or the duration of the incubation and infectious periods of the disease. Given the lack of quantitative behavioural data, particularly data concerning patterns of sexual contact between age classes, sexual-activity classes (risk groups) and between urban and rural population centres, it is not possible to make precise predictions about the likely demographic impact of AIDS over the coming decade in any one country.

Although model development has outstripped data availability, we believe a few general pointers emerge from recent theoretical work. Simple models based on random mixing between sexual activity classes and restricted mixing in age classes suggest that, with realistic assumptions concerning the major epidemiological and demographic parameters (such as doubling times in the general population between 1 and 3 years), AIDS is able to reverse the sign of population growth rates over timescales of a few to many decades^{27,44,45}. The added refinements of age-dependent changes in sexual behaviour, men having sexual contact with women younger than themselves and not infrequent male contact with a small proportion of women with high rates of sexual-partner change, all serve to accentuate the predicted demographic impact, with the only significant uncertainty being whether AIDS induced mortality will decrease population size over a few or many decades. We therefore continue to interpret the available facts as telling us that, in the absence of major changes in behaviour or the development and effective distribution of better drugs or a vaccine, AIDS is likely to induce significant demographic changes in some African countries. The fact that the disease already appears to be the leading cause of adult mortality in certain urban centres in Africa lends support to this conclusion.⁴

What can be done to reduce the spread of infection? A priority at present must be to acquire better data on patterns of sexual behaviour (both rates of partner change and mixing patterns), not only to improve interpretation and prediction, but, more important, to help target education and the distribution of condoms to the groups where they will have the most beneficial impact. For example, frequent use of condoms by female prostitutes and their male clients could delay or even prevent a more widely disseminated epidemic among the general population in countries where infection in high-risk groups is low to moderate at present (for example, Nigeria). In addition, the growing body of evidence that points to the importance of other sexually transmitted diseases (STDs) in enhancing the likelihood of transmission of HIV^{29,46}, argues for greatly increased efforts to treat and control STD spread in general. Simple models of concomitant STD and HIV transmission lend support to this approach, provided the degree to which the other disease enhances HIV transmission is sufficiently large⁴⁷. A further

advantage of greater effort in general STD control is that such programmes facilitate counselling and condom distribution to a high-risk segment of the population. Models can also be used to assess the likely influence on the rate of spread of infection of different degrees of behaviour change or frequency of condom use, in relation to the timing of the introduction of such behaviour changes in the course of the epidemic. The effects of timing are important, and not intuitively obvious, given the nonlinear character of the epidemic. But the general conclusion is that changes introduced earlier on in the course of the epidemic have a disproportionately greater effect than similar changes introduced later⁴⁸. Therefore, even when levels of HIV infection are low in the general population and AIDS is not the leading cause of mortality in countries afflicted by many other serious diseases, significant resources should be directed towards inducing behavioural change to try to prevent a widely disseminated lethal epidemic in 10–20 years time. Models that chart the slow but continuous development of the epidemic over many decades are important in convincing governments and international agencies to take this course, despite many more obvious and immediate public-health priorities.

Future epidemiological research needs are many but the immediate concerns must lie with the scientific assessment of different strategies of intervention, concomitant with quantitative studies of behaviour to assess sexual activity and mixing, both before and after a given intervention. It is vitally important that, in addition to measuring success through changes in rates of infection (which are difficult to interpret, given the nonlinear nature of epidemics), emphasis is placed on assessing precisely who changes behaviour. More generally, Tables 1 and 2 highlight the need for better longitudinal data on seroprevalence (ideally from cohorts), estimates of incubation and infectious periods, and studies of the probabilities of vertical and horizontal transmission, all in African settings.

Epidemiological research in Africa and elsewhere has tended to be on a small scale, with little emphasis on the standardization of study design and methods. The successes of multicentre cohort studies in the United States and Europe involving many patients suggest much could be gained in Africa by encouraging carefully designed multicountry studies using standard methods. Some progress has been made in this direction by the WHO in sexual-behaviour research, but more emphasis is likely to be of great benefit in other areas of the epidemiological study of AIDS. But the growing suspicion that genetic variability in viral populations, both in and between infected individuals, is a major influence on the likelihood of transmission and the development of disease will necessitate the addition of molecular virological techniques to the already difficult problems of studying transmission events that involve highly variable patterns of human sexual behaviour. □

R. M. Anderson, M. C. Boily, G. P. Garnett and J. T. Rowley are at the Parasite Epidemiology Research Group, Department of Biology, Imperial College, London University, London SW7 2BB, UK; R. M. May is at the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

- Gottlieb, M. S. *et al.* *New Engl. J. Med.* **305**, 1425–1431 (1981).
- Barre-Sinoussi, F. *et al.* *Science* **220**, 868–871 (1988).
- Gallo, R. C. *et al.* *Science* **224**, 497–500 (1984).
- De Cock, K. M. *et al.* *Science* **249**, 793–796 (1990).
- Ryder, R. W. *et al.* *New Engl. J. Med.* **320**, 1658–1662 (1989).
- Centers for Disease Control. *National HIV Seroprevalence Surveys 1–26* (US Department of Health and Human Services, 1990).
- Des Jarlais, D. C. *et al.* *J. Am. med. Ass.* **261**, 1008–1012 (1989).
- Vanichseni, S. *et al.* *V Int. Conf. AIDS Montreal 6/4–9 Poster TGO 23* (1989).
- John, J. T., Babu, G. P., Jayakumari, H. & Simoes, E. A. F. *Lancet* **i**, 160–161 (1989).
- Biggar, R. J. *AIDS* **4**, 1059–1066 (1990).
- Goodenow, M. *et al.* *J. AIDS* **2**, 344–352 (1989).
- Wellings, K. *et al.* *Nature* **348**, 276–278 (1990).
- Anzala, A. *et al.* *V Int. Conf. AIDS Montreal abstract A025* (1989).
- N'galy, B. & Ryder, R. W. *J. AIDS* **1**, 551–558 (1988).
- Rogers, M. F. *et al.* *Pediatrics* **79**, 1008–1014 (1987).
- Taylor, J. M. G., Kwo, J. M. & Detels, R. *J. AIDS* **4**, 69–75 (1991).
- Goodenow, M. *et al.* *J. AIDS* **2**, 344–352 (1989).

18. Nowak, M., May, R. M. & Anderson, R. M. *AIDS* **4**, 1095–1103 (1990).
19. Hira, S. K. *et al. Br. med. J.* **299**, 1250–1253 (1989).
20. Italian Multicentre Study *Lancet* **ii**, 1043–1045 (1988).
21. Blanche, S. *et al. New Engl. J. Med.* **320**, 1643–1648 (1989).
22. European Collaborative Study *Lancet* **i**, 253–259 (1991).
23. Scott, G. B. *et al. New Engl. J. Med.* **321**, 1791–1796 (1989).
24. Ryder, R. W. *et al. AIDS* **4**, 725–732 (1990).
25. Berkley, S. *et al. AIDS* **4**, 1237–1242 (1990).
26. Brown, R. C. *AIDS* **4**, 1267–1270 (1990).
27. Anderson, R. M., May, R. M. & McLean, A. R. *Nature* **332**, 191–290 (1988).
28. Levy, J. A. *AIDS* **4**, 1051–1058 (1990).
29. Piot, P. & Laga, M. *Br. med. J.* **298**, 623–624 (1989).
30. Konings, E., Anderson, R. M., Morley, D. & Meegan, M. *AIDS* **3**, 245–247 (1989).
31. Anderson, R. M., May, R. M., Ng, W., Rowley, J. T. *Phil. Trans. Roy. Soc. B* (in the press).
32. Carael, M., Cleland, J. & Odeokun, L. *AIDS* (in the press).
33. Hogsborg, M. & Aaby, P. in *Anthropological Studies Relevant to the Sexual Transmission of HIV* (International Union for the Scientific Study of Population, Liege, in the press).
34. Potts, M., Anderson, R. M. & Boileau, M. C. *Lancet* (in the press).
35. Anderson, R. M., Ng, W., Boily, M. C. & May, R. M. *N.Y. Acad. Sci.* **569**, 240–274 (1989).
36. Anderson, R. M. & May, R. M. *Nature* **333**, 514–522 (1988).
37. Hyman, J. M. & Stanley, E. A. *Math. Biosci.* **90**, 415–474 (1988).
38. Bongaarts, J. *Stats. Med.* **8**, 103–120 (1989).
39. Jacques, J. A., Simon, C. P., Koopman, J. *et al. Math. Biosci.* **92**, 119–199 (1988).
40. Gupta, S., Anderson, R. M. & May, R. M. *AIDS* **3**, 1–11 (1989).
41. Sharp, P. M. & Li, W. H. *Nature* **336**, 315 (1988).
42. Yokoyama, S., Chung, L. & Gojobori, T. *Molec. Biol. Evol.* **5**, 237–251 (1988).
43. May, R. M. & Anderson, R. M. *Parasitology* **100**, S89–S101 (1990).
44. May, R. M., Anderson, R. M. & McLean, A. R. *Math. Biosci.* **90**, 475–505 (1988).
45. May, R. M., Anderson, R. M. & McLean, A. R. *Lect. Notes Biomath.* **81**, 220–248 (1988).
46. Kreiss, J., Carael, M. & Mehous, A. *Genitourin. Med.* **64**, 1–2 (1988).
47. May, R. M. & Anderson, R. M. in *Current Topics in AIDS* Vol. 2 33–68 (Wiley, New York, 1989).
48. Rowley, J. T., Anderson, R. M. & Ng, T. W. *AIDS* **4**, 47–56 (1990).
49. *HIV/AIDS Surveillance Data Base* 9th update (US Bureau of the Census, Washington DC, 1990).
50. Andreasson, P. *et al. V Int. Conf. AIDS Montreal* 6/4–9 poster MGP 16 (1989).
51. Gnaore, E. *et al. VI Int. Conf. AIDS San Francisco* 6/20–24 poster FC663 (1990).
52. Mati, J. *et al. VI Int. Conf. AIDS San Francisco* 6/20–24 poster (1990).
53. Liomba, N. G. *et al. V Int. Conf. AIDS Montreal* 6/4–9 poster WGO 29 (1989).
54. Maiga, M. K. *et al. V Int. Conf. AIDS Montreal* 6/4–9 abstract A577 (1989).
55. *HIV Seroprevalence: Blood-Bank, Mozambique* (Ministry of Health (Mozambique), 1989).
56. Le Page, P. *et al. IV Int. Conf. AIDS and Associated Cancers in Africa* Marseille poster 243 (1989).
57. Killewo, J. *et al. AIDS* **4**, 1081–86 (1990).
58. Guay, L. *et al. VI Int. Conf. AIDS San Francisco* 6/20–24 poster (1990).
59. Murile, A. *V Int. Conf. AIDS Montreal* 6/4–9 abstract TBP 353 (1989).
60. Piot, P. *et al. J. infect. Dis.* **155**, 1108–1112 (1987).
61. Lifoon, A. R. *et al. VI Int. Conf. AIDS San Francisco* abstract C33 (1990).
62. Ward, J. *et al. IV International Conference on AIDS* Stockholm abstract 7711 (1988).
63. Giezecke, J. *et al. Br. med. J.* **297**, 99–102 (1988).
64. Rezza, G. *et al. AIDS* **3**, 87–90 (1989).
65. Auger, I. *et al. Nature* **336**, 575–577 (1988).
66. European Collaborative Study *Lancet* **ii**, 1039–1042 (1988).
67. Fischl, M. A. *et al. J. Am. med. Ass.* **257**, 640–644 (1987).
68. European Study Group *Br. med. J.* **298**, 411–415 (1989).
69. Padian, N. *et al. J. Am. med. Ass.* **258**, 788–791 (1987).
70. Glaser, J. B., Strange, T. J. & Rosati, D. *Archs intern. Med.* **149**, 645–649 (1989).
71. Johnson, A. M. *et al. AIDS* **3**, 367–372 (1989).
72. Peterman, T. A., Stoneburner, R. L., Allen, J. R., Jaffe, H. W. & Curran, J. W. *J. Am. med. Ass.* **259**, 55–58 (1988).
73. Hewlett, I. K. *et al. J. AIDS* **3**, 714–720 (1990).
74. France, A. J. *et al. Br. med. J.* **296**, 526–529 (1988).
75. Centres for Disease Control *Morbidity and Mortality Weekly Reports* **36**, 593–595 (1987).
76. Biberfeld, G. *et al. Scand. J. infect. Dis.* **18**, 497–500 (1986).
77. Andinan, W. A. *et al. AIDS* **4**, 758–766 (1990).
78. Goedert, J. J. *et al. Lancet* **ii**, 1351–1354 (1989).
79. Reeves, G. K. *Phil. Trans. R. Soc. B* **325**, 109–114 (1989).
80. Rothenberg, R. *et al. New Engl. J. Med.* **317**, 1297–1302 (1987).
81. Anderson, R. M. *J. statist. Soc. A* **151**, 66–93 (1988).
82. Blower, S. M., Anderson, R. M. & Wallace, P. J. *AIDS* **3**, 763–772 (1990).
83. Baldwin, J. D. & Baldwin, J. I. *J. Sex Res.* **1**, 181–196 (1988).
84. Hudson, C. P. *et al. AIDS* **2**, 255–260 (1988).
85. Van de Perre, P. *et al. Lancet* **ii**, 524–527 (1985).
86. Berkley, S. F. *et al. IV Int. Conf. AIDS and Associated Cancers in Africa* Marseille 1989 poster 012 (1989).
87. Melbye, M. *et al. Lancet* **ii**, 1113–1115 (1986).
88. Lisse, I. M. *et al. AIDS* **4**, 1263–1266 (1990).
89. Thangcharven, P. *et al. HIV Infection in Thailand* (Mahidol University Publication, Bangkok, 1989).
90. Ntozi, J. P. M. & Lubega, M. in *Anthropological Studies Relevant to the Sexual Transmission of HIV* (International Union for the Scientific Study of Population, Liege, in the press).
91. Boily, M. C. & Anderson, R. M. *IMA J. math. Med. Biol.* (in the press).

ACKNOWLEDGEMENTS. We thank P. Way (permission to use the US Bureau of Census HIV/AIDS surveillance data base), Wan Ng and H. Udy. This work was supported by the Overseas Development Administration, the MRC and the Royal Society (to R.M.M.).

Ancient oceans, ice sheets and the hydrological cycle on Mars

V. R. Baker*, R. G. Strom*, V. C. Gulick*†, J. S. Kargel*, G. Komatsu* & V. S. Kale‡

* Lunar and Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA

† Department of Geosciences, University of Arizona, Tucson, Arizona 85721, USA

‡ Department of Geography, University of Poona, Pune 411 007, India

A variety of anomalous geomorphological features on Mars can be explained by a conceptual scheme involving episodic ocean and ice-sheet formation. The formation of valley networks early in Mars' history is evidence for a long-term hydrological cycle, which may have been associated with the existence of a persistent ocean. Cataclysmic flooding, triggered by extensive Tharsis volcanism, subsequently led to repeated ocean formation and then dissipation on the northern plains, and associated glaciation in the southern highlands until relatively late in martian history.

MARS is the only planet besides Earth known to have had a long-term dynamic hydrological cycle. Despite its modern cold, dry climate and extensive areas of heavily cratered terrain left over from ancient times, Mars shows abundant evidence for the action of water and ice earlier in geological history. This includes fluvial dissection^{1,2}, periglacial landforms^{3,4}, ice-related

permafrost^{5,6}, lacustrine features^{7,8}, possible phased degradation of impact craters⁹ and possible glacial landforms¹⁰. The most spectacular fluvial features on the planet are the immense outflow channels, which originated by cataclysmic outbursts of water from subsurface sources^{11,12}. The largest outflow channels are concentrated in the Chryse trough on the eastern margin of the Tharsis bulge. The channels nearly all debouched their flood water to the northern plains of Mars, where ponded sediments^{13,14}, shoreline indicators, infilled craters and basins, spillover channels¹⁵, pitted basin floors, whorled patterns of multiple ridges and other evidence indicate the past existence of extensive lakes¹⁵ or temporary regional flooding¹⁶.

Parker *et al.*¹⁷ presented evidence for the sporadic formation of a great ocean covering the northern plains of Mars, which we name Oceanus Borealis. Features related to the inundation of the northern plains and the outflow channels all belong to the Hesperian and Amazonian geological systems¹⁸. These time-stratigraphic designations (the Hesperian is the older, the Amazonian the younger system) both occur after the period of heavy bombardment, corresponding to the Noachian geological system¹⁸, which ended $\sim 2.8\text{--}3.8 \times 10^9$ years ago¹⁹. Here we propose a conceptual scheme in which Oceanus Borealis repeatedly forms and dissipates, temporarily producing a relatively warm,

wet climate. The short episodes of maritime climate explain otherwise anomalous landform assemblages, most notably the extensive development of Amazonian-age glacial landforms in the southern hemisphere²⁰, south of about -45° . The widespread distribution of these glaciated terrains²¹ documents the development and retreat of a massive polar ice cap, which we name the Austral ice sheet, extending from the south pole to $\sim 45^{\circ}$ S, and in places to $\sim 33^{\circ}$ S. The glacial ages of at least three regions (Argyre, Hellas and Dorsa Argentea) are Middle Amazonian²¹.

In each episode of inundation, the flood water on the northern plains was gradually lost to the associated maritime climate by infiltration into the global regolith, thereby replenishing groundwater and ground-ice reserves, and by deposition as ice sheets in the high-latitude regions. Eventually much of the atmospheric CO_2 precipitated, returning the planet to the cold, dry climate and tenuous CO_2 atmosphere that characterizes it today. Thus, Mars has experienced long periods of slow degradation²², punctuated by exceptional, short periods of concentrated geomorphological activity.

Fluvial history

The earliest geological processes inferred from martian landforms¹⁸ are intense impact cratering and widespread volcanism, which occurred in the Noachian time-stratigraphic system. During this period, fluvial valley networks developed extensively in the ancient heavily cratered terrains of Mars^{23,24}. A possible erosional episode of these ancient cratered terrains occurred near the end of the period⁹. Ground-water sapping, a process in which ground-water outflow undermines slopes²³, was appropriately concentrated to form most of the Noachian valleys^{24,25}. This process can form valleys only if long-term hydraulic head differentials are maintained, so that they can drive water flow in subsurface aquifers^{26,27}. A flow persistent enough for valley erosion cannot be achieved by simple melting of an ice-rich permafrost²⁸. A prolonged flow of ground-water

that could produce sapping could be driven by hydrothermal water circulation in saturated permeable rock. Such circulation systems may have developed near impact²⁹ or volcanic³⁰ sites, but would have been localized. The widespread distribution of Noachian martian valleys seems to require either the planet-wide repetition of such local endogenetic heating, or an atmospheric cause, resembling the hydrological cycling on Earth. Given the very high impact rates and widespread volcanism during the Noachian, either mechanism could produce widespread valley development by persistent runoff to an ocean, evaporation and precipitation over upland regions. The very high porosities of common martian surface rocks (lava flows and impact-brecciated regolith) would allow subsurface aquifers to be replenished easily, so that head differentials could be sustained and drive prolonged ground-water flow. The resulting sapping³¹ would then produce the observed pattern of structurally controlled, low-density valley networks²⁴.

Although Noachian valleys are consistent with atmospheric water cycling, including precipitation, and direct evidence for the associated oceanic sink would have been removed by subsequent erosion and deposition. Possible water-storage basins, including the northern plains and the Argyre and Hellas impact basins, are filled with post-Noachian rocks¹⁸. The existence of a Noachian ocean is not essential, as the valley networks can be explained by hydrothermal circulation with minimal atmospheric effects. It is, however, a theoretical consequence of a warm, wet early Mars^{32,33}, and it is a probable element in the hydrological cycling responsible for the widespread valley networks described above. Most valley development seems to have ceased at the end of the late heavy bombardment²⁴, indicating that the persistent ocean, if it existed, disappeared at this time. The timing of the transition is modelled³² as the evolution of a relatively dense primordial CO_2 atmosphere that maintained the appropriate climate until the end of the Noachian (end of late heavy bombardment). The subsequent martian climate is

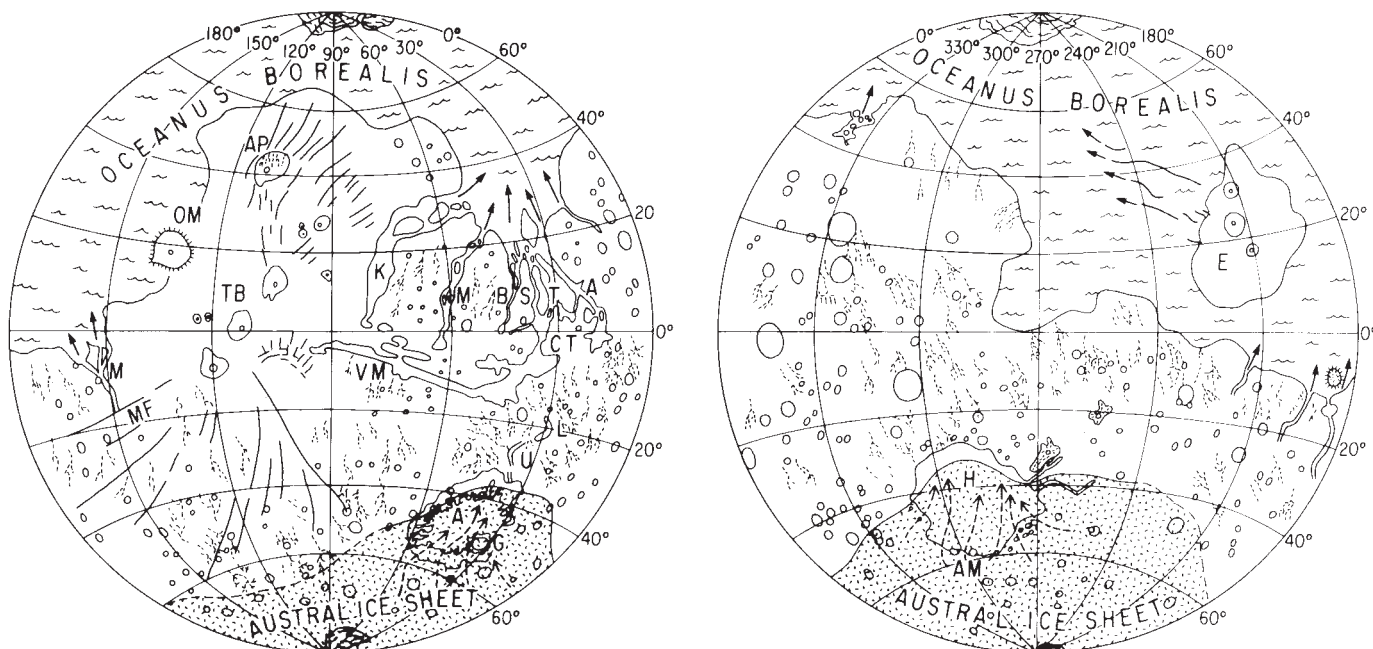


FIG. 1 Sketch maps of western (left) and eastern (right) hemispheres of Mars showing the past extent of northern plains occupied by Oceanus Borealis (wave pattern), areas modified by the austral ice sheet (stippled pattern) and flow directions indicated by outflow channel landforms (solid arrows) and by the assemblage of glacial landforms (dashed arrows). Additional features discussed in the text include the Tharsis bulge (TB) with associated linear fracture systems such as Memnonia Fossae (MF) and the

Valles Marineris (VM); the Chryse trough (CT); the volcanic complexes Elysium (E), Olympus Mons (OM) and Alba Patera (AP); the impact basins Argyre (A) and Hellas (H); Amphrites plateau (AM); and the Galle impact crater (G). Outflow channels include Uzboi (U), Ladon (L), Mangala (M), Kasei (K), Maja (J), Shalbatana (B), Simud (S), Tiu (T) and Ares (R). The networks of dashed lines in the heavily cratered upland terrains show the general pattern of dissection by fluvial valleys.

TABLE 1 Hypsometry of Mars northern plains (Oceanus Borealis)

Contour* (km)	Volume (10^7 km^3)	Area (10^7 km^2)	Average depth (km)	Equiv. water layer† (m)
0	6.5	3.8	1.7	450
-1.0	3.1	2.8	1.1	210
-2.0	1.0	1.4	0.7	70

* Water below stated contour.

† Covering entire surface.

commonly assumed to have been too cold and dry to allow a dynamical hydrological cycle to take place. However, recent work by J. F. Kasting (personal communication) suggests that although the faintness of the early Sun may have prevented a warm early Mars, increasing solar luminosity later in the planet's history permits a warmer climate and denser atmosphere. As our work centres on temporary changes in climate associated with post-Noachian ocean formation, our hypothesis does not require the presence or absence of an earlier period of warming.

Some martian valley formation continued after the heavy bombardment had ceased^{26,34,35}. The post-Noachian valleys indicate that either these valleys formed without extensive atmospheric water cycling (by hydrothermal circulation), or the conventional wisdom of a continuous cold, dry climate during the Hesperian to Amazonian is wrong. Valleys of Amazonian age are developed on a mantle of ash, probably of low permeability, on the northern flanks of Alba Patera volcano^{26,35}. The Alba valleys are the martian landforms most similar to terrestrial stream valleys formed by the result of rainfall, but they are too young to be consistent with the view that warm, wet conditions on Mars terminated completely after late heavy bombardment. The cumulative amount of water required to erode the valleys on Alba Patera, estimated from analogous Hawaiian valleys, is equivalent to a column of water ~250 km high in 10^5 years^{35,36}. The immense amount of water that must be cycled is remarkably similar to that required for formation of the Amazonian-age glacial features mentioned above. The time required to produce the assemblage of posulated martian glacial features formed by the Austral ice sheet is estimated from terrestrial rates to be 10^5 to 2×10^7 years. Comparable terrestrial glacial landscapes require the cycling of a cumulative amount of ice equivalent to a 200-km-high column in $\sim 2 \times 10^6$ years²¹. Both glaciers and valleys can be explained by dynamic cycling of water in the atmosphere and the associated surface-water ponding during the Amazonian age.

Formation and dissipation of oceans

The less-common post-Noachian martian valley networks and widespread glacial features were formed at approximately the same time as subsurface water was catastrophically released to form the outflow channels. An explanation of their origin is suggested by the observed association of the largest outflow channels with great fracture zones³⁷ extending from the Tharsis bulge (Fig. 1). The fractures, it is suggested, served as immense flow conduits, the Memnonia Fossae conducting the flow west from Tharsis to Mangala Vallis³⁸, and the Valles Marineris system and parallel fractures conveying the flow eastward to the Kasei and Maja Valles (valleys?) as well as to the channels of the Chryse trough (the Shalbatana, Simud, Tiu and Area valleys). Together these channels account for well over 90% of the outflow discharges. Although the ages of the channels span the Hesperian and Amazonian³⁹, there is abundant evidence that outflow has occurred several times from individual channels⁴⁰. Many channelling events may have been simultaneous within the uncertainties of the measurements.

After late heavy bombardment, the style of valley development

TABLE 2 Filling rates for Oceanus Borealis

	0	Contour (km) -1.0	-2.0
Fill rates:			
$10^9 \text{ m}^3 \text{ s}^{-1}$	2 yr	1 yr	15 wk
$10^{10} \text{ m}^3 \text{ s}^{-1}$	11 wk	5 wk	10 d
Source zone thickness at:			
25% porosity	2.6 km	1.3 km	0.4 km
10% porosity	6.6 km	3.2 km	1.0 km

Based on water volume below stated contour.

changed from a more widespread occurrence in the ancient cratered terrains to a highly localized nature on the flanks of several volcanoes. This change also coincided with the change from extensive volcanism in the cratered terrains to volcanism concentrated at Tharsis and Elysium. If water was involved in atmospheric cycling from Noachian ocean, then it became locked in the permeable upper martian crust, comprising an ice-rich permafrost of maximum thickness perhaps 1–3 km (ref. 6), overlying liquid ground water. The prevailing post-Noachian martian climate became cold and dry, much as it is today, but disrupted briefly at intervals.

The disruptive trigger was volcanism, and the accompanying plutonism and tectonism. The Tharsis bulge and its associated volcanic and fracture systems probably resulted from doming and rifting of the lithosphere over a massive mantle plume. Possible terrestrial analogues might be mantle plumes associated with terrestrial flood-basalts and rifted margin volcanism or, on a larger scale, 'superplume' activity⁴¹. As modelled by White and McKenzie⁴², terrestrial flood-basalt events were concentrated in short periods ($\sim 10^6$ yr) after an anomalously hot mantle gradually built up to a critical threshold for decompressive melting. Assuming similar conditions for Tharsis volcanism, then massive, rapid emplacement of magma (we will assume $\sim 10^7 \text{ km}^3$ to be affected⁴²) into the martian lithosphere would have: (1) injected $\sim 3 \times 10^{32}$ erg of thermal energy, sharply increasing the heat flux (from a Mars average of $\sim 25 \text{ erg cm}^{-2} \text{ s}^{-1}$ to $\sim 50 \text{ erg cm}^{-2} \text{ s}^{-1}$, if the anomalous magmatic heat is released over 3×10^6 yr) and the geothermal gradient (from 25 to 50 K km^{-1}); (2) melted huge amounts of stored ground ice ($\sim 5 \times 10^6 \text{ km}^3$ for 5% conversion of thermal energy into latent heat of fusion of ice); (3) set up an enormous hydrothermal circulation system drawing ground water from adjacent regions toward the thermal anomaly; and (4) driven the water through very permeable volcanic rocks to the great fracture systems and out onto the surface. Eruption of ground water by thermally driven convection could be assisted by violent exsolution of CO_2 . Sudden destructive discharge of water, forming outflow channels and chaotic terrain, could occur where the permafrost cap thinned to some critical threshold thickness (on average the melting geotherm in Tharsis would rise from a depth of 3 km to 1.5 km) or where fractures associated with magmatic intrusions propagated through the ground-water zone to the surface.

If we use appropriate hydraulic formulae¹, the dimensions and gradients of the outflow channels indicate total combined discharge rates of 10^9 – $10^{10} \text{ m}^3 \text{ s}^{-1}$. This total requires individual channel peak discharges of 10^7 – $10^9 \text{ m}^3 \text{ s}^{-1}$ for flow depths of 100–500 m and maximum velocities of 8–60 m s^{-1} . The immense flows required to form the channels⁴³ imply exceptionally high hydraulic conductivities probably generated by liquefaction of extensive subsurface zones of water-saturated, porous materials³⁷, such as unconsolidated impact breccia. The subsurface liquefaction would have resulted in collapse to form chaotic terrains and various troughs of Valles Marineris.

Such catastrophic flood discharges to the northern plains of Mars would have enormous consequences. Table 1 summarizes

the possible volumes of standing water. These figures are approximate because of large uncertainties (as great as ± 1.5 km) in the existing topographic data⁴⁴. The present geometry of the northern plains could accommodate bodies of water up to oceanic proportions, as much as 4×10^7 km² in area, holding up to 6.5×10^7 km³ water and having an average depth up to 1,700 m. The maximum water volume that might reside in Oceanus Borealis is equivalent to a planetwide water layer 450 m deep. This figure is consistent with other estimates of the martian water inventory⁴⁵.

The outflow channels have a complex history of multiple flooding events over a prolonged period⁴⁰. We suggest that at intervals, simultaneous discharge from the channels was triggered by the planetary-scale volcanism described above. The consequences of such episodes are summarized in Table 2. At total combined discharge rates consistent with the indicated outflow channel dimensions, the various sizes of Oceanus Borealis determined in Table 1 would be achieved in periods of days to a few years. The source zone of martian upland terrain that could provide this water transfer would have to be saturated with water and ice to some thickness. The volumes indicated for various sizes of Oceanus Borealis require saturated thicknesses averaging between 0.5 and 6.6 km over the source region, depending upon the assumed porosity. The elevated region of Tharsis, underlain by permeable lava flows, would probably provide unusually thick source regions from which to derive outflow discharges.

Climate implications

The massive transfers of water through the outflow channels, from subsurface storage to Oceanus Borealis, would generate profound changes in climate. Figure 2 illustrates the main flows of mass and energy. The outpouring of subsurface water would rapidly release dissolved CO₂ and water vapour to the atmosphere. Additional large amounts of these two greenhouse gases would be generated by inundation of the polar carbon dioxide cap, and by evaporation of water and sublimation of ice in the newly formed Oceanus Borealis. As martian temperatures rose

TABLE 3 Possible sources of CO₂ related to catastrophic outflow and ocean formation on Mars

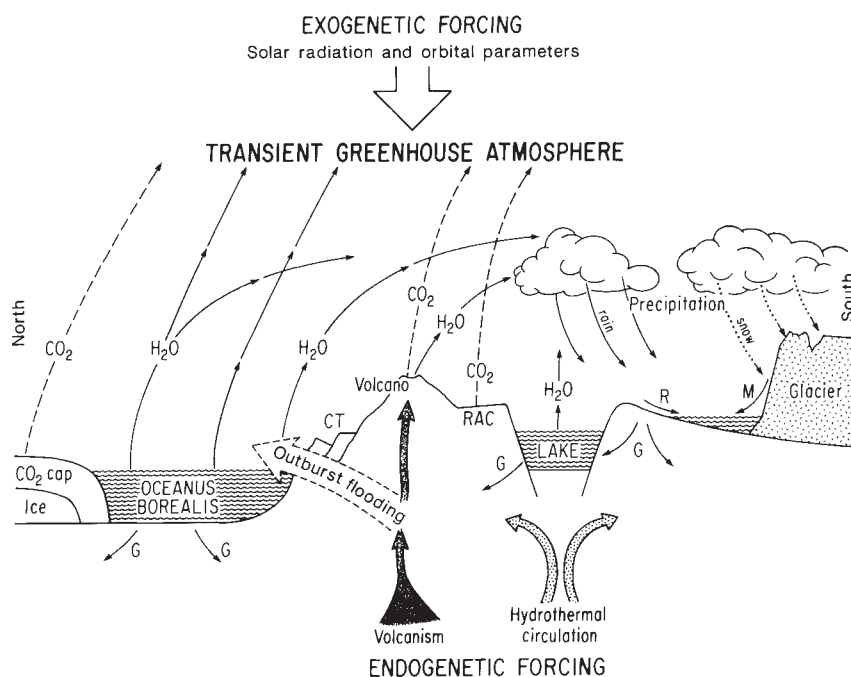
Source	CO ₂ partial pressure (mbar)
North polar cap	~20
Massive volcanism	~100
Adsorbed on regolith: ocean basin	≤100
land	≤1,300
CO ₂ clathrate	≤4,000

during the resulting transient greenhouse conditions, additional global change could occur by melting ground ice and perhaps CO₂ clathrate in the permafrost of the heavily cratered uplands, and possibly by releasing adsorbed CO₂ from the previously cold regolith⁴⁶. The planet would experience a period of maritime climate with exogenetic hydrological cycling from ocean to atmosphere to land by precipitation. Elevated regions next to the ocean, such as Alba Patera, would receive abundant orographic precipitation^{26,35}. Colder regions, such as the high elevations of the south polar latitudes, would receive snow, resulting in glacial activity²¹. Low- to mid-latitude regions could have a very arid climate, with little effect on the geomorphology. The climate and atmosphere-surface interactions would be modulated by periodic changes of orbital eccentricity and obliquity⁴⁷.

Sources of CO₂

Carbon dioxide is an essential greenhouse gas for driving the modification of martian climate. Table 3 lists the sources and estimated amounts of carbon dioxide. At the greatly reduced solar luminosity of the earliest history of Mars, Pollack and others³² calculate that an atmospheric pressure of ~2–5 bar is required to bring the annual average martian surface temperature to the melting point of water ice. Less is needed for summertime equatorial melting. Even less atmospheric CO₂,

FIG. 2 Schematic representation of ocean-land-atmosphere interactions associated with the formation of Oceanus Borealis on Mars. Volcanism and associated heat flow (heavy arrows) cause the hydrothermal circulation of groundwater and trigger outburst flooding from reservoirs of ground ice and water in the highly permeable upper martian crust. In the outburst flooding, immense discharges from a collapsing, liquifying aquifer system result in chaotic terrain (CT). The huge flux of water delivered to the northern plains releases water vapour (through evaporation) and carbon dioxide gas (by warm water inundation of the polar CO₂ cap and CO₂-rich regolith, and through exsolution from the released ground water). Warmed by these radiatively active gases, the resulting 'greenhouse' atmosphere precipitates snow to form glaciers in the south polar regions, eventually making up the austral ice sheet. The temporary maritime climate might also supply rain, recharging groundwater flow systems (G) and sustaining temporary lakes. Additional hydrological activity includes glacial meltwater (M) and runoff (R) of overland flow at local areas of low porosity. Warming might also release CO₂ adsorbed on clay particles in the highlands regolith (RAC), forming an additional positive feedback to the climate, although initiated from within the planet, this hydrological cycle would be modulated by the exogenetic forcing of solar radiation and changes of martian orbital parameters. Unlike on Earth, the ocean-related cycling did not function continuously through martian history. Sporadic hydrological activity ended when recharged water became trapped as ground ice or water in the upper martian



crust and when carbon dioxide became trapped as carbonates. For the activity to be repeated the cycle of endogenetic processes would have to be renewed and free carbon dioxide would have to be available.

perhaps 0.5–1.0 bar, is needed to produce widespread warm, wet conditions late in Mars history because the solar luminosity is close to that of today. The possible sources of carbon dioxide from massive, cataclysmic volcanism and water outflow are sufficient to produce these amounts (Table 3). Furthermore, feedback mechanisms involving the atmospheric effects of water vapour will add greatly to surface warming.

Igneous activity in the Tharsis region could emit up to 100 mbar of CO₂. This estimate assumes an extruded and intruded magma volume equivalent to a layer of lava 2 km thick spread over the Tharsis plains, and further assumes the magma initially contained 0.65 wt% CO₂. This amount of dissolved CO₂ is similar to the amount in Hawaiian magma⁴⁸, and is appropriate for saturation at a pressure of 13 kbar (equivalent to a depth of ~120 km on Mars).

Massive igneous activity could elevate regional geotherms and destabilize any CO₂ clathrate in the regolith⁴⁹, thus releasing additional CO₂ to the atmosphere. We estimate an upper limit on the amount of released gas by assuming that the entire ocean was formed from melting clathrate rather than pure ice. Using the maximum ocean from Table 1, and assuming the stoichiometry CO₂·5³H₂O and 60% cage occupancy, we calculate that over 4 bar of CO₂ would be released along with the water. Although this calculation suggests that clathrate may be a very large source of CO₂, we do not expect that all the water in the martian regolith actually would be in the form of clathrate; hence, 4 bar is an upper limit.

Another source of CO₂ could be ground water confined below the permafrost zone (1–3 km). Although terrestrial ground water typically contains concentrations of up to 0.02 wt% CO₂, under high pressures it can store much larger amounts of dissolved CO₂. On Mars, subsurface water may contain additional CO₂ derived over geological time from degassing of magma at depth or from basal melting of a clathrate permafrost. To place an upper limit on the total amount of CO₂ that could be dissolved in ground water, we assumed that all water involved in ocean formation was fully saturated with CO₂ at an average hydrostatic pressure of 300 bar and a temperature of 12 °C. From the data tabulated in Linke⁵⁰, this is approximately 8 wt%. Given the

maximum ocean size in Table 1, then ~1.3 bar of CO₂ can be dissolved in the ground water. If the pressure is suddenly released by geological processes, then exsolution of the gas phase could catastrophically drive out the confined water from subsurface aquifers.

Once water is released, inundating and heating the cold regolith of the northern plains, much of the CO₂ physically adsorbed in the regolith would be liberated. We estimate that as much as 100 mbar of gas could be liberated from the regolith of the northern plains alone (area ~4 × 10⁷ km²), assuming 500 m of regolith with an average mass fraction of CO₂ in the regolith equal to 0.009 (the average value for basalt, nontronite and palagonite⁵¹), and assuming present atmospheric conditions. If these sources of gases affect the climate enough to warm the planet significantly, additional adsorbed CO₂ in highlands regolith could be released, potentially amounting to an additional several hundred millibars. But the net amount of gas that would be released is uncertain, as there is a negative feedback associated with increasing atmospheric pressure which tends to drive more gas back into the regolith.

As indicated by multiple phases of Oceanus Borealis¹⁷ and by the ranges of ages of the valley networks²⁶, the above activity must have stopped, then been reactivated after long periods of hydrological inactivity. The termination of a cycle may have been aided by seeping of water into martian surface materials. Except at local sites of volcanic ash or sedimentary mantling²⁶, any precipitated water would tend to infiltrate rapidly. As more water became sequestered into the martian uplands, the polar ocean would shrink, reducing its influence on the atmosphere. Other negative feedbacks might include a planetary albedo increase from extended ice and snow cover, thereby lowering the mean surface temperature. As the maritime climate phase ended, the bulk of the martian water would return to storage as ground ice and water, persisting as such until another phase of thermal activity. Episodes of volcanism at Tharsis or Elysium probably led to the ocean forming repeatedly, with effects on the climate and geomorphology that depended on the scale of the ocean. Over the history of the planet, the declining internal heat flow⁵², atmospheric escape of hydrogen⁵³ and sequestering

TABLE 4 Generalized hydrological history of Mars

System ¹⁸	Cratering rate	Volcanism	Valley networks	Outflow channels	Oceans	Glacial and periglacial activity	Inferred climate
Amazonian	Low	Concentration at Tharsis and Elysium	Local high-density networks on Alba Patera ^{26,35}	Considerable outburst flooding during Early and Middle ³⁸ ; Lesser flooding during Late ^{15,54}	Transient and episodic during Early and Middle; Late formation of small seas and lakes ¹⁵	Extensive Middle glacial of southern hemisphere ²¹ ; Late formation of periglacial debris mantles ^{3,4,8} and polar layered deposits	Cold and dry punctuated by episodes of wetter glacial and periglacial conditions
Hesperian	Transitional	Concentration at volcanic plains; initiation of Tharsis and Elysium	Local development on mantling units ³⁴ and intercrater plains ²⁴	Major phases of outburst flooding in late Hesperian ¹⁸	Transient, episodic and more extensive than in the Amazonian ¹⁷	Possible ancient phases of glaciation at Argyre, Hellas and Dorsa Argentea ²¹	Cold and dry punctuated by relatively warm, wet pluvial conditions
Noachian	High	Widespread highland volcanism	Widespread dissection of the heavily cratered uplands ^{23,24}	Not significant	Possible persistent ocean inferred from models ^{32,33}	Not known	Possible warm, wet pluvial conditions ³²

Continuing loss of water probably occurred by exospheric escape⁵³ and hydrolysis of silicate minerals, contributing to reduced hydrological activity, which also coincided with declining internal heat flow⁵², declining cratering rates and localization of volcanism. Oxygen released by dissociation of water probably contributed to extensive oxidation of the regolith.

of CO₂ in carbonates³² and H₂O in clays would gradually reduce the activity in successive cycles. The youngest discharges of subsurface water produced relatively small outflow channels postdating the lava flows on the Tharsis plains⁵⁴ and Elysium plains units^{15,39}. These outflow channels may have formed within the last few hundred million years, and the associated effects on the climate were probably small.

Implications

Table 4 is a summary of our preliminary interpretation of the ocean and the interactions with land and atmosphere throughout martian history. Our conceptual scheme involves both a possible progressive change from warm, wet ancient conditions to the cold, dry modern climate, and episodes of short-term, warmer and wetter climate associated with cataclysmic formation of oceans during later martian history. Following this hypothesis, diverse observations can be related to an evolutionary mechanism of global planetary change. Whether one is sympathetic or antagonistic to the scheme, work aiming either to confirm or reject it should help elucidate the causes of the immense range of martian geomorphological features. In accord with Conant's

philosophical formulation⁵⁵, we hope that our conceptual scheme leads to further experimentation and observations. Numerous deductions from the hypothesized interactions between ocean, land and atmosphere could be effectively explored with predictive climate models. The stratigraphic correlation between different landforms, and their relative ages, should be more precisely established.

A principal uncertainty that still remains is whether the various ocean phases persisted long enough for the generation of life. There are strong arguments that photosynthetic life on Earth would have been destroyed by vaporization of much of the ocean water in basin-forming impacts until 3.8 Gyr BP (the end of late heavy bombardment)⁵⁶. As photosynthetic life was apparently present at 3.56 Gyr BP (ref. 57) and possibly by 3.77 Gyr BP (ref. 58), it seems that life may have developed and evolved rapidly, becoming capable of photosynthesis in less than 10⁷ or 10⁸ years. For Mars, any of the active hydrologic phases could have led to the inception and further development of life^{59,60}. The most likely place for fossils of any primitive martian life to be preserved is in the marine or lacustrine sediments of an active hydrological phase. □

Received 22 April; accepted 9 July 1991.

1. Baker, V. R. *The Channels of Mars* (University of Texas Press, Austin, 1982).
2. Mars Channel Working Group. *Geol. Soc. Am. Bull.* **94**, 1035–1054 (1983).
3. Carr, M. H. & Schaber, G. G. *J. geophys. Res.* **82**, 4039–4054 (1977).
4. Squyres, S. W. *J. geophys. Res.* **84**, 8087–8096 (1979).
5. Lucchitta, B. K. *Icarus* **45**, 264–303 (1981).
6. Rossbacher, L. A. & Judson, S. *Icarus* **45**, 39–59 (1981).
7. Nedel, S. S., Squyres, S. W. & Anderson, D. W. *Icarus* **70**, 409–441 (1987).
8. Squyres, S. W. *Icarus* **79**, 229–288 (1989).
9. Chapman, C. R. & Jones, K. L. *Rev. Earth planet. Sci.* **5**, 515–540 (1977).
10. Lucchitta, B. K. *J. geophys. Res.* **87**, 9951–9973 (1982).
11. Baker, V. R. & Milton, D. J. *Icarus* **23**, 27–41 (1974).
12. Carr, M. H. *J. geophys. Res.* **84**, 2995–3007 (1979).
13. Jöns, H. P. *17th Lunar planet. Sci. Conf.* 404–405 (Lunar and Planetary Institute, Houston, 1985).
14. Lucchitta, B. K. *18th Lunar planet. Sci. Conf.* 498–499 (Lunar and Planetary Institute, Houston, 1986).
15. Scott, D. H. & Chapman, M. G. *Proc. Lunar planet. Sci.* **21**, (eds Ryder G & Sharpston V. L.) 669–677 (Lunar and Planetary Institute, 1991).
16. Lucchitta, B. K., Ferguson, H. M. & Summers, C. A. *NASA Tech. Memo.* 88383, 450–453 (National Aeronautics and Space Administration, Washington D C, 1985).
17. Parker, T. J., Saunders, R. S. & Schneeberger, D. M. *Icarus* **82**, 111–145 (1989).
18. Scott, D. H. & Tanaka, K. L. *US Geol. Surv. Geologic Series Map I-1802-A*, scale 1:15,000,000 (US Geological Survey, Denver, 1986).
19. Strom, R. G., Croft, S. K. & Barlow, N. G. *Mars* (University of Arizona Press, in the press).
20. Kargel, J. S. & Strom, R. G. *21st Lunar planet. Sci. Conf.* 597–598 (Lunar and Planetary Institute, Houston, 1990).
21. Kargel, J. S. & Strom, R. G. *Science* (submitted).
22. Arvidson, R. E., Guinness, E. A. & Lee, S. W. *Nature* **278**, 533–535 (1979).
23. Carr, M. H. & Clow, G. D. *Icarus* **48**, 91–117 (1981).
24. Baker, V. R. & Partridge, J. B. *J. geophys. Res.* **91**, 3561–3572 (1986).
25. Pieri, D. C. *Science* **210**, 895–897 (1980).
26. Gulick, V. C. & Baker, V. R. *J. geophys. Res.* **95**, 14325–14344 (1990).
27. Howard, A. D. *NASA Tech. Memo.* 4210, 342–344 (National Aeronautics and Space Administration, Washington DC 1990).
28. Goldspiel, J. & Squyres, S. W. *Icarus* **89**, 392–410 (1991).
29. Brakenridge, G. R., Newsom, H. E. & Baker, V. R. *Geology* **13**, 859–862 (1985).
30. Gulick, V. C., Marley, M. S. & Baker, V. R. *19th Lunar planet. Sci. Conf.* 441–442 (Lunar and Planetary Institute, Houston, 1988).
31. Baker, V. R., Kochel, R. C., Laity, J. E. & Howard, A. D. in *Ground-water Geomorphology* (eds Higgins, C. G. & Coates, D. R.) 235–266 (Spec. Pap. 252, Geological Society of America, Boulder, Colorado, 1990).
32. Pollack, J. B., Kasting, J. F., Richardson, S. M. & Poliakov, K. *Icarus* **71**, 203–224 (1987).
33. Schaefer, M. W. *J. geophys. Res.* **95**, 14291–14300 (1990).
34. Grant, J. A. III Providence, Rhode Island, 1990).
35. Gulick, V. C. & Baker, V. R. *Nature* **341**, 514–516 (1989).
36. Gulick, V. C., Marley, M. S., Baker, V. R. *22nd Lunar planet. Sci. Conf.* 509–510 (Lunar and Planetary Institute, Houston, 1991).
37. MacKinnon, D. J. & Tanaka, D. L. *J. geophys. Res.* **94**, 17359–17370 (1989).
38. Tanaka, K. L. & Chapman, M. G. *J. geophys. Res.* **95**, 14315–14323 (1990).
39. Tanaka, K. L. *J. geophys. Res.* **91**, E139–E158, (1986).
40. Rotto, S. L. & Tanaka, K. L. *22nd Lunar planet. Sci. Conf.* 1135–1136 (Lunar and Planetary Institute, Houston, 1991).
41. Larson, R. L. *Geology* **19**, 547–550 (1991).
42. White, R. S. & McKenzie, D. P. *J. geophys. Res.* **94**, 7685–7730 (1989).
43. Robinson, M. S. & Tanaka, K. L. *Geology* **18**, 902–905 (1990).
44. US Geological Survey *Topographic Maps of the Western, Eastern Equatorial and Polar Regions of Mars*, Map I-2030 (1989).
45. Carr, M. H. *Nature* **326**, 30–34 (1987).
46. Fanale, F. P. & Cannon, W. A. *J. geophys. Res.* **84**, 8404–8414 (1979).
47. Bills, B. G. *J. geophys. Res.* **95**, 14137–14153 (1990).
48. Bottinga, Y. & Javoy, M. *Bull. Volcanol.* **53**, 73–85 (1991).
49. Milton, D. J. *Science* **183**, 654–656 (1974).
50. Linke, W. F. *Solubilities: Inorganic and Metal-Organic Compounds A-Ir*, Vol. 1, 461 (D. Van Nostrand Co. Inc., Princeton, 1958).
51. Zent, A. P., Fanale, F. P., & Postawko, S. E. *Icarus* **72**, 241–249 (1987).
52. Schubert, G. & Spahn, T. *J. geophys. Res.* **95**, 14095–14104 (1990).
53. Hunten, D. M., Donahue, T. M., Walter, J. C. G. & Kasting, J. F. in *Origin and Evolution of Planetary Atmospheres* (eds. Atreya, S. K., Pollack, J. B. & Matthews, M. S.) 386–422 (University of Arizona Press, Tucson, 1989).
54. Mouginis-Mark, P. J. *Icarus* **84**, 362–373 (1990).
55. Conant, J. B. *Science and Common Sense* (Yale University Press, 1951).
56. Sleep, N. H., Zahnle, K. J., Kasting, J. F. & Morowitz, H. J. *Nature* **342**, 139–142 (1989).
57. Schopf, S. W. & Walter, M. R. in *Earth's Earliest Biosphere* (ed. Schopf, S. W.) 214–239 (Princeton University Press, 1983).
58. Schidlowski, M. *Nature* **333**, 313–318 (1988).
59. McKay, C. P. *Adv. Space Res.* **6**, 269–285 (1986).
60. McKay, C. P. *Icarus* **91**, 93–100 (1991).

ACKNOWLEDGEMENTS. We thank S. K. Croft and T. J. Parker for helpful comments. We also acknowledge the late H. Faul, who suggested oceans and glaciers on Mars in an unpublished 1973 paper. This research was supported by the NASA, Planetary Geology and Geophysics Program.

Common West African HLA antigens are associated with protection from severe malaria

Adrian V. S. Hill, Catherine E. M. Allsopp, Dominic Kwiatkowski*, Nicholas M. Anstey*, Patrick Twumasi*, Pamela A. Rowe†, Stephen Bennett‡, David Brewster‡, Andrew J. McMichael & Brian M. Greenwood*

Molecular Immunology Group, Institute of Molecular Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK

* Medical Research Council Laboratories, Fajara, The Gambia

† Department of Epidemiology and Population Sciences, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK

‡ Royal Victoria Hospital, Banjul, The Gambia

A large case-control study of malaria in West African children shows that a human leucocyte class I antigen (HLA-Bw53) and an HLA class II haplotype (DRB1*1302-DQB1*0501), common in West Africans but rare in other racial groups, are independently associated with protection from severe malaria. In this population they account for as great a reduction in disease incidence as the sickle-cell haemoglobin variant. These data support the hypothesis that the extraordinary polymorphism of major histocompatibility complex genes has evolved primarily through natural selection by infectious pathogens.

THE genes of the major histocompatibility complex (MHC) which encode the human leucocyte antigens (HLA) are the most polymorphic known in man (Fig. 1)¹. Although it was proposed that most of this variation is essentially neutral^{2,3}, several observations now indicate that MHC polymorphism must be maintained by some type of selection pressure. Over the millions of years in which HLA genes have been evolving^{4,5}, most of the current alleles should have been lost by genetic drift unless some form of balancing selection was operating⁶, and the unusually even distribution of HLA frequencies in human populations, without a predominance of any single allele, also signifies balancing selection pressures^{7,8}. Furthermore, mutations which produce amino-acid substitutions cluster around the peptide-binding cleft of HLA molecules⁹⁻¹¹ indicating that selection must have been acting on this site at some stage during evolutionary history. However, such analyses cannot show that selection is ongoing today, nor determine whether it acted through differential viability, differential fertility or some other process.

The discovery of MHC restriction of the immune response, whereby foreign antigens could only be recognized by T lymphocytes when presented in association with class I or class II MHC molecules, led to the proposal that MHC polymorphism is maintained by different alleles providing varying degrees of protection against infectious pathogens¹². An individual who is heterozygous for MHC alleles would have an increased capacity to present antigens from a range of pathogens compared with

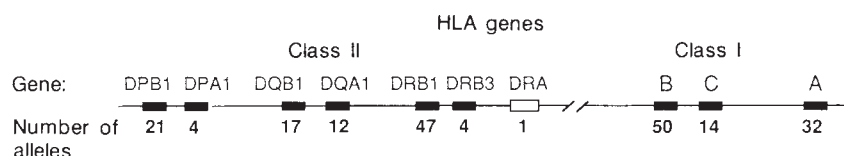
homozygotes. This theory of pathogen-driven MHC diversity requires that individuals of different MHC types should differ in their susceptibility to at least some major infectious pathogens. However, no convincing associations between HLA polymorphism and susceptibility to commonly fatal infectious diseases have yet been identified, in contrast to the well-known associations between HLA and many autoimmune diseases¹³. This has led to the proposal of many alternative theories to explain the existence of MHC polymorphism in man and other species¹⁴⁻¹⁶.

We report a case-control study of the influence of HLA class I and II variation on susceptibility to *Plasmodium falciparum* malaria, a disease which has been a major selective force in recent human evolution¹⁷, and in which T-lymphocyte dependent mechanisms are important in providing protective immunity¹⁸⁻²². By detailed HLA typing of over 2,000 individuals from the West African population of The Gambia we show that an HLA class I antigen, HLA-Bw53, and an HLA class II haplotype, DRB1*1302-DQB1*0501, are associated with reduced susceptibility to severe malaria. Interestingly, both of these HLA types are much more frequent in West Africans than other racial groups, arguing that natural selection by malaria has contributed to their elevated frequencies in Africa. These data support the hypothesis that MHC polymorphism is maintained primarily by altering susceptibility to infectious pathogens.

Malaria in West Africa

P. falciparum malaria is a major cause of childhood mortality, currently accounting for between one half and two million deaths per year in sub-Saharan Africa alone²³. In The Gambia, about 1% of children under 5 years old die each year from malaria²⁴. It is clear that although infection by *P. falciparum* is extremely common in African children only a small proportion of these infections lead to life-threatening illness²⁴. Many factors are likely to influence the extent of disease progression including genetically determined differences in host susceptibility. The best known host resistance gene is that encoding the β -chain of the haemoglobin (Hb) variant, HbS, which provides a striking illustration of natural selection by malaria²⁵. No comparable data exist on the relevance of HLA antigen variation, although two limited studies^{26,27} conducted outside Africa have suggested that HLA class I alleles would be worth more detailed investigation. Piazza *et al.* noted some increased variation for certain

FIG. 1 Map of the location of polymorphic HLA class I and class II genes in the human major histocompatibility complex on chromosome 6. The number of known alleles of each gene, whether identified serologically or by sequence analysis, is indicated^{1,44}.



HLA types between highland and lowland populations exposed in the past to different malarial endemicities in Sardinia²⁸ but, as the authors and others²⁹ point out, there are several alternative explanations for these findings.

It is clear from work on the HbS polymorphism and susceptibility to malaria that the strongest protection is from severe malaria and death with less protection from mild illness and very little from parasitaemia²⁵. To increase the probability of

TABLE 1 Gambian severe malaria study

	Number	Median age	Hb AS (%)
Severe malaria	619	3.0	1.2
severe malarial anaemia	221	1.8	
cerebral malaria	429	3.6	
Mild controls	510	1.9	12.9
Mild malaria	354	3.0	2.8
Severe controls	332	2.2	10.9
Healthy adults	220	28	16.8

Number of individuals recruited in each clinical category, median age and HbS carrier (=Hb AS genotype) rate. The HbAS percentages are based on results from 1,802 of the 2,035 individuals (S. Abdalla *et al.*, manuscript in preparation). The carrier rate in the children with severe malaria (1.2%) was significantly lower ($\chi^2=52.5$, $P<10^{-11}$) than in the mild controls (12.9%). The adults had a higher HbAS frequency than both the mild ($\chi^2=1.1$) and severe ($\chi^2=3.4$, $P=0.07$) control children, which is only in part explained by the small expected increase in carrier rate with age as a result of differential mortality from malaria²⁵. **Design:** Children up to the age of 10 years were recruited at the Royal Victoria Hospital, Banjul, and the Medical Research Council Hospital, Fajara, 8 miles away, during the months of maximum malaria transmission³⁰ from August to November in 1988 and 1989, and also during August and September of 1990. Approval for the study was obtained from the Gambian government/MRC joint ethical committee and permission for venesection from a parent or guardian. A diagnosis of malaria was made if a child with an appropriate clinical picture had a parasitaemia $>2,500 \mu\text{l}^{-1}$, a level likely to be associated with illness in this setting, and relevant clinical and laboratory investigations excluded other diagnoses. Malaria cases were assigned to the severe group^{31,68,69} if any of the following features were present: (1) cerebral malaria, unrousable coma of score <3 (by standard criteria⁶⁹) or repeated prolonged (>30 min) seizures; (2) severe malarial anaemia, haemoglobin level $<5 \text{ g dl}^{-1}$ on admission; (3) hypoglycaemia, blood glucose level $<2.2 \text{ mmol}$. Fifty-six children had both cerebral malaria and severe malarial anaemia and 25 had hypoglycaemia alone. Of the 88 children who died, 72 had cerebral malaria, four had severe malarial anaemia and 12 had both: a further 33 were discharged with residual neurological deficits. All children fulfilling the operational definition of severe malarial anaemia had malaria, and several factors indicate that this was the major cause of their anaemia despite the low background haemoglobin levels of normal children in The Gambia⁶⁸. These are their high spleen and gametocyte rates, normochromic normocytic blood films, their very low HbS carrier rate (sickle cell homozygotes were excluded), and the striking concordance of the seasonality of this condition with that of cerebral malaria in The Gambia³⁰. Hence, although occasional cases may have had additional factors contributing to the anaemia, malaria was the major cause of the life-threatening state. As expected^{69,70}, parasite densities (geometric mean with 95% ci $\times 10^3 \mu\text{l}^{-1}$) correlated weakly with severity: cerebral malaria (84.9, 69.5–103.7), mild malaria (70.5, 60.8–81.8), and in the more chronically infected severe malarial anaemia group (61.5, 39.6–95.4). Although parasite densities were higher among those who died (122.4, 85.2–175.8) than survivors (70.7, 62.7–79.6), the fairly weak correlation between parasite density and clinical severity probably relates to numerous factors including the higher rate of prior antimalarial administration among the severe cases⁷⁰, the greater importance of sequestered than peripheral blood parasites³¹ and the complex dynamics of *P. falciparum* parasitaemias⁷¹. The cases of severe malaria were matched, as a group, to the other three groups of children for two variables: age and area of residence around Banjul. The 'mild controls' (see text) were recruited at both hospitals and at health centres in the study area. 'Severe controls' were inpatients at the two hospitals with a large range of other acute, mainly infectious, illnesses, but without evidence of current or recent malarial infection. These controls were considered separately from the mild (outpatient) controls because (ref. 72 and above data) they may be more susceptible to malaria than the general population, as found elsewhere in Africa⁷²: this effect may arise from an increased likelihood of children who are susceptible to frequent malarial episodes becoming severely ill during other infections, for example because of malaria-induced immunosuppression. Children with unexplained anaemia and known HbS homozygotes were excluded as controls. Two HbSS individuals were found among 603 severe malaria cases and 6 of 254 mild controls analysed were found to be previously unsuspected HbSS homozygotes. The ethnic composition of the population in this area is mixed: Mandinka (42%), Jola (14%), Wolof (14%), Fula (12%) and several less common ethnic groups. The children were not matched for ethnic group: instead analyses stratified for this variable were done. HLA allele frequency differences between the ethnic groups were small and the proportions of each ethnic group among the clinical categories were found to be very similar. The healthy adults were male blood donors matched retrospectively to the children for ethnic group.

detecting a significant effect of HLA variation we studied children with very severe illness, the majority of whom would have died without treatment. Most malarial deaths in African children are attributable to severe malarial anaemia or cerebral malaria^{30,31}. Although the precise pathophysiological mechanisms of these two major complications are unknown, it appears that severe malarial anaemia is often associated with failure to clear parasites from the circulation so that continuing red cell destruction leads to severe anaemia, whereas cerebral malaria is associated with blockage of cerebral blood vessels by parasitized erythrocytes³². We therefore consider each of these conditions separately, particularly for the analysis of HLA class II antigens which, through their influence on antibody formation and hence parasite clearance³³, might be expected to have a larger effect on severe malarial anaemia.

From 1988–90 we studied more than 600 cases of severe malaria (Table 1), severity being defined by standard clinical criteria which are powerfully predictive of outcome, at least for cerebral malaria. Children with severe malarial anaemia had a malarial illness with an extremely low and life-threatening haemoglobin level ($<5 \text{ g dl}^{-1}$). Hence most of these children, like those with cerebral malaria, were likely to have died without treatment even though, because of the effectiveness of transfusion therapy, few of the 88 hospital deaths were among this group. The severe malaria cases were compared with a large group of control children, frequency matched to the cases for age and area of residence. These 'mild controls' were children with mild, mostly infectious, illnesses who did not require hospital admission and did not have any malarial parasites in their blood on microscopy. Samples were also collected from three further groups to assist in the interpretation of any association seen in the primary comparison and to address particular issues. These were children with 'mild malaria', healthy adult blood donors and children hospitalized with illnesses other than malaria, 'severe controls', as defined in Table 1.

Determination of the carrier rates for HbS was performed to provide a measure against which any protective effect of HLA types could be assessed. The expected low frequency of carriers of HbS was found in the children with severe malaria (1.2%) compared with the mild controls (12.9%, $\chi^2=52.5$, $P<1 \times 10^{-11}$), the relative risk (RR) of 0.08 (95% confidence intervals (ci) 0.04–0.16) indicating that children who are carriers of HbS have more than 90% protection from severe malaria (Table 1).

HLA class I antigens

HLA class I antigens present peptides to cytotoxic T cells which, in certain mouse models of malaria, are induced by sporozoite infection and provide effective immunity by acting against the liver stage of the malaria parasite^{19–22}. To look for an association between HLA class I antigens and susceptibility to malaria in man we adopted a two-stage strategy. Initially we determined the HLA class I types of roughly half the cases of severe malaria and controls serologically (Table 2). Although both HLA-A24 and HLA-B14 were more common among the cases of severe malaria, their frequencies were low and the differences were of marginal statistical significance. The frequency of HLA-Bw53, a common antigen in this population, was significantly decreased (15.7%) among the cases of severe malaria compared with either the mild controls (24.3%, $\chi^2=4.3$, $P=0.04$), all the control children (22.9%, $\chi^2=4.7$, $P=0.03$), or the healthy adults (25.0%, $\chi^2=4.2$, $P=0.04$). The decrease in frequency was apparent both for cases of severe malarial anaemia (14.7%) and of cerebral malaria (16.1%).

However, because in this serological study we have tested for an association between susceptibility to malaria and 45 different HLA class I antigens, a single HLA association may have turned up simply by chance as a result of making numerous comparisons. So, as stage two, we retested the association on the remaining unanalysed samples asking a single question: is the frequency of HLA-Bw53 lower among the cases of severe

TABLE 2 HLA class I antigen frequencies

	Severe malaria <i>n</i> =306		Mild controls <i>n</i> =144		Severe controls <i>n</i> =171		Healthy adults <i>n</i> =112	
	Number	%	Number	%	Number	%	Number	%
A1	41	13.7	13	9.0	22	12.9	17	15.2
A2	74	24.2	41	28.5	46	26.9	32	28.6
A3	27	8.8	8	5.6	21	12.3	7	6.3
A23	85	27.8	32	22.2	43	25.1	26	23.1
A24	6	2.0	0	0	0	0	0	0
A26	44	14.4	25	17.4	26	15.2	13	11.6
A28	53	17.3	35	24.3	34	19.9	16	14.3
A29	12	3.9	7	4.9	5	2.9	3	2.7
A30	73	23.9	29	20.1	34	19.9	31	27.7
A31	1	0.3	2	1.4	1	0.6	1	0.9
A32	24	7.8	8	5.6	8	4.7	10	8.9
Aw33	67	21.9	33	22.9	39	22.8	29	25.9
Aw34	11	3.6	6	4.2	5	2.9	3	2.7
Aw36	1	0.3	0	0	0	0	1	0.9
B7	41	13.4	17	11.8	20	11.7	17	15.2
B8	67	21.9	31	21.5	26	15.2	18	16.1
B13	3	1.0	2	1.4	1	0.6	0	0
B14	30	9.8	5	3.5	14	8.2	4	3.6
B15	6	2.0	7	4.5	4	2.3	0	0
B17	68	22.2	27	18.8	39	22.8	24	21.4
B418	20	6.5	7	4.9	6	3.5	6	5.4
B22	9	2.9	2	1.4	8	4.7	2	1.8
B27	13	4.2	5	3.5	2	1.2	3	2.7
B35	86	28.1	48	33.3	53	31.0	32	28.6
B37	2	0.7	1	0.7	0	0	3	2.7
B39	6	2.0	4	2.8	6	3.5	2	1.8
B40	5	1.6	1	0.7	1	0.6	2	1.8
Bw41	9	2.7	4	2.8	6	3.5	2	1.8
Bw42	16	5.2	5	3.5	7	4.1	7	6.3
B44	12	3.9	1	0.7	8	4.7	4	3.6
B45	13	4.2	5	3.5	3	1.8	3	2.7
B49	25	8.2	13	9.0	14	8.2	11	9.8
Bw50	14	4.6	4	2.8	10	5.8	5	4.5
B51	17	5.6	7	4.9	7	4.1	4	3.6
Bw52	2	0.7	1	0.7	1	0.6	0	0
Bw53	48	15.7	35	24.3	37	21.6	28	25.0
Bw70	46	15.0	22	15.2	22	12.9	19	17.0
Cw1	15	4.9	8	5.6	9	5.3	6	5.4
Cw2	50	16.4	17	11.8	21	12.3	24	21.4
Cw3	91	29.8	52	36.1	54	31.6	33	29.5
Cw4	80	26.2	40	27.8	52	30.4	33	29.5
Cw5	20	6.6	7	4.9	6	3.5	6	5.4
Cw6	43	14.1	23	16.0	23	13.5	16	14.3
Cw7	55	18.0	17	11.8	25	14.6	19	17.0
Cw8	26	8.5	5	3.5	11	6.5	4	3.6

Number of individuals with each HLA class I antigen and the percentage antigen frequencies in the groups defined in Table 1. Significant frequency differences for individual antigens are described in the text. Subdivision of individuals with the HLA-Bw53 antigen into heterozygotes and apparent homozygotes showed the following numbers of the latter in each category: severe malaria, 11; mild controls, 6; severe controls, 5; adults, 4. Thus, there is no evidence of a significant difference in the relative risks of severe malaria between heterozygotes and apparent homozygotes for HLA-Bw53. The frequencies of individuals apparently homozygous for HLA-A, -B and -C alleles were also compared in these groups. A significantly lower frequency of HLA-B homozygotes was found in the severe malaria group, 17.3%, compared with the pooled control children, 25.5% ($\chi^2=5.7$, $P=0.017$; M-H for ethnic group $\chi^2=5.9$, $P=0.015$, $RR=0.60$ (0.40–0.91)). No significant difference was found in HLA-A or -C homozygosity rates although the latter are particularly affected by a high frequency of serologically undetected alleles. **Statistical analysis:** In all tables this was done using the χ^2 test with continuity corrections. Allowance for confounding factors was made by the Mantel-Haenszel test and by logistic regression⁵⁶. Since severe malaria is comparatively rare, the relative risk (RR) may be estimated by the odds ratio; this is given with an associated 95% confidence interval. The problem of multiple comparisons was dealt with for HLA class I antigens by a two-stage strategy, and for class II haplotypes by multiplying each *P*-value by the number of concurrent tests performed. Serological types were determined by the standard microlymphocytotoxicity assay using 180 well-characterized antisera on fresh or cryopreserved cells. From the total number of children entered into the study those for serological typing were chosen randomly, limited only by the availability of sufficient viable lymphocytes. More severe controls were included than mild controls because only tiny volumes of blood were available from most of the latter. The following specificities were tested for but not found: HLA-A11, -A25, -Aw43, -B38, -Bw47, -Bw48 and -Bw73.

malaria than in the mild controls? To do this, we cloned and sequenced the HLA-Bw53 allele^{34,35} and devised a method of typing for this sequence using allele-group-specific amplification by the polymerase chain reaction (PCR) and oligonucleotide hybridization. Using the PCR method to test the remaining samples (Table 3), the frequency of HLA-Bw53 was 16.9% among the cases of severe malaria and 25.4% in the mild controls ($\chi^2=6.4$, $P=0.008$; Mantel-Haenszel test (M-H) for ethnic group $\chi^2=6.4$, $P=0.01$, $RR=0.61$), confirming the association suggested by the serological results. Combining the serological and PCR data the relative risk of severe malaria among individuals with HLA-Bw53 compared with those without this allele is 0.59 (0.43–0.81). We also used the PCR method to type the remaining adult controls who again showed a frequency of about 25%, and the cases of mild malaria who had a lower, but not significantly different, frequency of 22.6% (Table 3).

The HLA-Bw53 antigen is present at highest frequency in sub-Saharan Africa: in 40% of Nigerians³⁶, 25% of Gambians, 21% of Zambians³⁷ and 16% of Zimbabweans³⁸, but is much less frequent in black South Africans (2%)³⁹. It is rare in caucasian and oriental populations (0–1%)^{40,41} and apparently absent in Pacific islanders⁴² and Amerindians⁴¹. Hence, this HLA class I allele, which is associated with significant protection from severe malaria, has many similarities in its global distribution to the classical protective haemoglobin variant, HbS (ref. 25).

HLA class II haplotypes

The importance of antibodies in defence against the blood stage of malaria infection has been demonstrated in several ways^{18,33,43}. So, to look for an HLA class II association with

TABLE 3 HLA-Bw53 association with severe malaria

	Serology		PCR	
	No.	HLA-Bw53 (%)	No.	HLA-Bw53 (%)
Severe malaria	306	15.7	307	16.9
Mild controls	144	24.3	364	25.4
Mild malaria	—	—	353	22.6
Healthy adults	112	25.0	106	26.4

HLA-Bw53 antigen frequencies determined by serology (Table 2) and by PCR. There is no overlap between the samples analysed by the two methods. **Analysis of interactions:** Of the seven individuals heterozygous for HbS among the severe malaria cases, 1 (14.3%) had the HLA-Bw53 antigen; 97 (16.6%) of 586 cases of severe malaria with the Hb AA genotype had this antigen. Analysis of a possible interaction between HLA-Bw53 and the HLA class II haplotype, DRw13.02-DQw1, associated with protection from severe malarial anaemia (Table 4), showed that, of the 34 individuals with DRw13.02 in the severe malarial anaemia group (out of 216 typed for both antigens), 4 (11.8%) had HLA-Bw53. Of the 138 mild controls with DRw13.02 (out of 477 typed for both antigens), 39 (28.3%) had HLA-Bw53; *RR* of severe malarial anaemia in individuals with both antigens compared to individuals with neither = 0.21 (0.05–0.60), which is similar to the product of the individual antigen relative risks. Similar analysis of the HLA-Bw53/DRw13.02 interaction among all severe malaria cases also showed no significant interaction. Hence, although the numbers with both antigens are relatively small, there does not appear to be any strong epistatic interaction between these alleles. **Analysis of parasite densities:** Because of numerous confounding factors, such as the high rate of prior treatment particularly among severe cases⁷⁰, this study could only have detected very large effects of genetic factors on peripheral blood parasite density: to avoid increasing the number of comparisons assessed, and thereby reducing the power of the study, this supplementary analysis was not part of the study design. Although the parasite density (geometric mean with 95% CI $\times 10^3 \mu\text{L}^{-1}$) among malaria cases (studied within 2 h of arrival at hospital, *n*=523) with HLA-Bw53 (72.6, 55.9–94.3) was lower than in those without this antigen (77.1, 67.7–87.8) neither this difference (Wilcoxon test: $P=0.76$) nor the larger reduction in mean parasite density among those with HLA-DRw13.02 (68.7, 54.2–87.1) compared with those without (81.1, 71.0–92.6; $P=0.23$) was statistically significant. PCR typing was done for HLA-Bw53 as described elsewhere⁷³. Briefly, the primers 5'-CCGGAACACACAGATCTT-3' and 5'-GTGCTAGCGGACTGGTC-3' were used to specifically amplify only HLA-Bw53 and HLA-B35 in a PCR buffer containing 10% DMSO and 1 mM MgCl₂, with 35 cycles of denaturation at 94 °C (90 s), annealing at 58 °C (120 s) and extension at 72 °C (120 s). HLA-Bw53 alleles were distinguished from HLA-B35 by hybridization with an oligonucleotide probe 5'-CGGATCGCGCTCCGCTAC-3' to the 3' end of exon 2, with washing in 6 × SSC at 59 °C, after immobilization of the amplification product on nitrocellulose filters.

susceptibility to the two forms of severe malaria, severe malarial anaemia and cerebral malaria, we determined HLA class II haplotypes defined by restriction fragment length polymorphisms (RFLP) of the DRB and DQB genes, a molecular method of HLA typing that has some advantages over alternative techniques. Analysis (Table 4) showed a highly significant difference in frequency for a single haplotype, with the X-I RFLP pattern, between the children with severe malarial anaemia and the mild controls. This haplotype pattern is invariably associated with the DRw13 and DQw1 serological specificities and subtyping by DNA sequencing and T cell assays has shown the DRw13 allele to be of the DRB1*1302 subtype⁴⁴⁻⁴⁸. This was confirmed for this Gambian population by sequencing of the DRB1 gene and typing of many further samples by allele-group-specific PCR and oligonucleotide hybridization⁴⁹ (Table 4). Like HLA-Bw53, this haplotype is less frequent among the cases of severe malarial anaemia (8.7% gametic frequency) than in the mild controls (16.4%, $\chi^2_1 = 14.0$, $P = 0.00018$, $P_{\text{corrected}} = 0.0023$), indicating that it is a protective rather than a susceptibility haplotype. Statistically significant protection was seen only from severe malarial anaemia, even though the frequency of this haplotype was also decreased in the children with cerebral and mild malaria compared with the mild controls. Analysis of genotype frequencies showed that, for each form of severe malaria, homozygotes for this haplotype had lower relative risks than heterozygotes (Table 4).

In almost all cases reported previously, the DRB1*1302 allele has been linked to the DRB3*0301 allele⁵⁰ and to DQ genes specifying the DQw6 (rather than the DQw5) subtype of DQw1 (ref. 51). However, sequencing of the DQA1 and DQB1 genes from a Gambian with a DRB1*1302 haplotype showed a novel arrangement of allelic sequences (Fig. 2): a DQB1*0501 rather than the expected DQB1*0604 sequence was linked to the usual DQA1 sequence⁵². Furthermore, this DRw13-DQw5 haplotype is more common in The Gambia than the caucasian DRw13-DQw6 haplotype (Fig. 2). Comparison of the protective efficacies of the two haplotypes against severe malarial anaemia showed that the associated relative risks were similar, suggesting that both are protective and that the DRB1*1302 and DRB3*0301 alleles that they have in common may be more important than the DQB gene for protection from severe malarial anaemia. However, although the more common DQw5

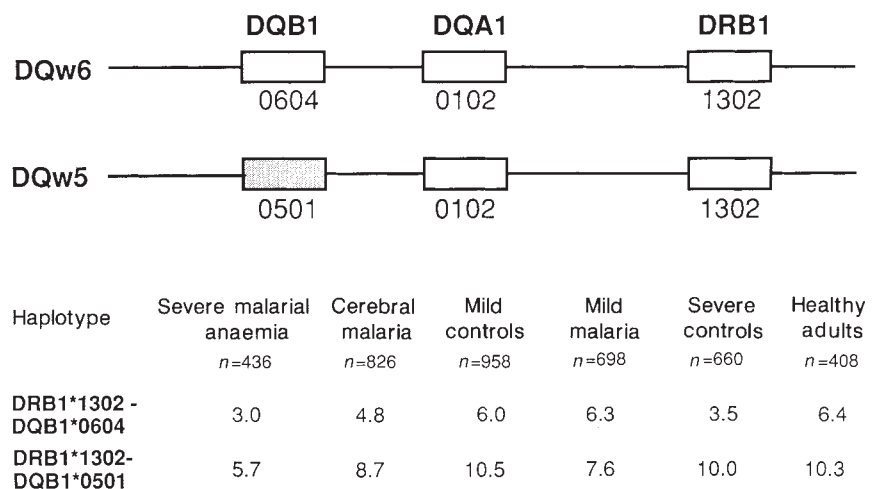
haplotype is clearly associated with protection, the less frequent DQw6 haplotype is difficult to assess with certainty, precluding a definite assignment of the association to particular DR or DQ genes. The protective DQw5 haplotype is probably the same as the DRw13-DQw5 haplotype found in American blacks^{45,48} and an Israeli cell line⁵³, but this haplotype is clearly very rare in northern Europeans⁵⁴ and north American caucasians⁵⁵. There was no significant⁵⁶ linkage disequilibrium between HLA-Bw53 and -DRB1*1302 in the population as a whole ($\Delta = 0.001$, $P > 0.7$), nor within clinical groups, indicating independent associations. Although this study was not designed to measure the influence of HLA antigens on peripheral blood parasite densities, retrospective examination of the available data showed no large differences (see Table 3). Analysis of the relative risks of severe malaria among individuals with more than one of the protective alleles, HbS, HLA-Bw53 and DRB1*1302 (see legends to Tables 2-4), showed no clear evidence of epistatic interactions, but only substantial effects would have been detectable.

HLA and protective immunity

The simplest explanation for these associations, amongst several possibilities, is that they relate to the antigen-presenting ability of HLA molecules. Thus, the association between HLA-Bw53 and protection from severe disease would suggest that, as in animal models¹⁹⁻²², class I-restricted T lymphocytes, presumably cytotoxic T cells (CTL), play an important part in providing protective immunity to malaria in man. Because HLA class I molecules are absent on human red blood cells, these CTL would probably act against the liver-stage malaria parasite. Similarly, a plausible explanation for the HLA class II association is that only individuals with this haplotype are able to present a particular conserved epitope on a blood-stage parasite antigen to helper T cells, leading to more rapid parasite clearance. It may be surprising that an infectious organism as large as a malaria parasite should give rise to a disease that is HLA-associated: *P. falciparum* must contain thousands of potential T-cell epitopes which will have varying HLA restriction patterns, and immune responses against many *P. falciparum* epitopes are indeed evoked⁵⁷. The existence of an HLA association implies that most such responses must be of limited protective value, and that the number of epitopes to which a strongly protective

FIG. 2 Top, Gene maps of the two DRB1*1302 haplotypes found in The Gambia. Bottom, Percentage frequencies of the two DRB1*1302 haplotypes in the various groups. The frequencies of these two haplotypes sum to the frequency of the DRw13.02-DQw1 haplotype given in Table 4. The DQw5 haplotype was significantly less frequent among the cases of severe malarial anaemia compared with the mild controls ($\chi^2_1 = 7.9$, $P = 0.005$, $RR = 0.52$ (0.32-0.82)) and the severe controls ($\chi^2_1 = 5.6$, $P = 0.017$). The DQw6 haplotype was significantly less frequent among the severe malarial anaemics compared with the mild controls ($\chi^2_1 = 4.9$, $P = 0.026$, $RR = 0.49$ (0.25-0.93)) but not compared with the severe controls ($\chi^2_1 = 0.1$). All 96 Gambian haplotypes with DRB1*1302, analysed by PCR and oligonucleotide hybridization⁴⁹, had the DRB3*0301 sequence⁵⁰ (corresponding to the serological specificity DRw52c), as is true in other populations.

METHODS. The DRB1, DQA1 and DQB1 genes were analysed from an individual homozygous for the DRw13.02 RFLP pattern (Table 4) by PCR using locus-specific primers⁵³ for exon 2, cloning by blunt-ended ligation into *Sma*I-cut m13mp8 vectors (Amersham) and dideoxy-sequencing. To type the DQB allele of all individuals with DRB1*1302, either a PCR or a Southern blot method was used. In the PCR method the second exon of the DQB genes was amplified using the primers 5'-CGCAGAGGATTTCTGTACAG-3' and 5'-GGGCGACGACGCTCA-CCTC-3' and hybridization was with oligonucleotides specific for DQB1*0501



(5'-CCGGGAGTGACGC-3') or DQB1*0604 (5'-GAGGTGGGTACCGC-3'). The latter will also hybridize with the uncommon DQB1*0605 allele⁴⁵ which, if present, will be included with *0604 in this comparison. The Southern blot method employed the restriction enzyme *Bam*HI and hybridization with a DQB probe which gave 7.5-kb and 3.4-kb bands with the DQw6 and DQw5 haplotypes, respectively⁴⁸.

immune response can be mounted may be small, a consideration of some importance for vaccine development.

Analysis of the mechanisms of the HLA associations described here may be helpful in identifying the antigens against which protective immune responses are directed. Although several *P. falciparum* antigens have been advocated or used as candidate vaccines, progress in vaccine development has been retarded by

our ignorance of the nature of protective immunity in man. A search for an association between the ability to respond to a particular malaria antigen or epitope and the presence of protective HLA types may be useful in identifying one of these antigens as a target of natural protective immunity and thus a preferred antigen for vaccine trials.

Evolutionary implications

This study provides the clearest evidence yet obtained that a lethal infectious pathogen, one which has had enormous effects on human history¹⁷, is influencing the evolution of polymorphic MHC genes in man. Although a study of survivors of typhoid and yellow fever epidemics in Surinam⁵⁸ suggested that HLA antigens may have influenced outcome in one or other of these diseases, only in Marek's disease of chickens⁵⁹ had there been a convincing association with a fatal natural infection. These African data indicate that pathogen-driven selection is operating in present-day human populations and that it is unnecessary to propose² that this phenomenon has been episodic with long intervening periods of selective neutrality. Another implication is that HLA associations with many other potentially fatal infectious diseases may exist, but that very large sample sizes, with detailed molecular typing, may be necessary to demonstrate these convincingly.

The degree of protection afforded to children with the HLA types we have identified allows some estimates to be made of the time it would take for these alleles to reach their present-day frequencies under natural selection by malaria. Assuming simple directional selection this would take between 2,000 and 7,000 years for HLA-Bw53, depending on the initial allele frequency assumed⁶⁰: estimates of the length of time *P. falciparum* malaria has been a major cause of mortality in Africa are similar¹⁷. However, we feel that selection pressures are unlikely to be this constant, and fluctuating selection pressures⁶¹⁻⁶³ due to epidemics of infectious diseases might actually help to maintain MHC diversity. In this regard, two other models of how MHC diversity could be maintained by differential protection from infectious pathogens have been advocated: overdominant^{6,12} selection (heterozygote advantage) and frequency-dependent selection^{64,65}, in which individuals with rare alleles have the greatest resistance to pathogens. It is a prediction of the overdominance model that heterozygotes, who should be fitter than homozygotes, would be less common in the disease group than among controls. However, among cases of severe malarial anaemia, HLA class II haplotype heterozygotes were significantly more frequent than in the controls (Table 4), as were heterozygotes for HLA-B alleles in the severe malaria group compared with controls (Table 2). Furthermore, homozygotes for DRB1*1302 had lower relative risks than heterozygotes both for severe malarial anaemia and for cerebral malaria, which may also be more compatible with frequency-dependent than overdominant selection. Although the precision of this analysis is limited by the inability of the typing methods used to detect all true heterozygotes, and other infectious pathogens may show different HLA associations, these data do not support the recent tendency to assume that balancing selection for MHC genes must be overdominant¹¹.

The importance of these HLA associations is emphasized by comparing the proportion of potential cases of severe malaria in the population which are prevented by the presence of HLA-Bw53, DRB1*1302 or HbS. Although HbS is clearly the most protective to an individual, and associated with a 92% reduction in the relative risk of severe malaria, only 13% of the Gambian population are carriers, so the proportion of potential cases prevented^{66,67} is 12%. HLA-Bw53 and the DRB1*1302 haplotype are less protective but they are much more prevalent, so that the proportion of potential cases prevented can be calculated to be 10.3% of severe malaria cases (HLA-Bw53) and 16% of cases of severe malarial anaemia (DRB1*1302): a total of about 15% of all potential severe malaria cases in The Gambia.

TABLE 4 HLA class II haplotypes

DR	Serotype DQ	RFLP pattern		Severe malarial anaemia <i>n</i> =436	Cerebral malaria <i>n</i> =826	Mild controls <i>n</i> =958	Mild malaria <i>n</i> =698	Severe controls <i>n</i> =660	Healthy adults <i>n</i> =408
		DR	DQ						
13.04	7	XXI	V	24.8	26.4	25.3	30.4	27.1	23.8
13.02	1	X	I	8.7	13.5	16.4	13.9	13.5	16.7
10	1	III	I	7.1	9.1	8.7	8.0	9.7	7.4
9	2	XVII	IX	7.8	7.1	7.7	5.9	4.9	6.4
4	8	XII	IV	5.7	4.4	5.2	4.6	4.4	4.7
1	1	I	I	5.1	5.1	4.3	2.7	3.6	3.4
11.01	7	XX	V	2.8	2.5	3.4	3.9	3.5	3.2
13.01	1	IX	II	3.4	2.8	3.4	3.6	3.6	1.7
8	7	V	V	3.2	2.7	4.0	3.4	3.0	3.0
3.01	2	VIII	III	3.4	2.8	2.2	2.2	3.0	1.5
7	2	XIV	VI	4.6	3.1	1.8	3.0	2.6	3.2
13.03	7	XV	V	2.5	2.4	2.0	2.0	1.8	2.9
3.02	4	XV	IV	1.6	2.2	1.0	2.4	2.9	2.2

Percentage frequencies of the 13 common HLA class II haplotypes in the various groups. The total number of haplotypes typed in each group (*n*) is shown (that is, twice the number of individuals). Haplotypes are identified by RFLP labels in Roman numerals⁴⁶ and the corresponding serological types are shown with the subtype of the DRw3, DRw11 and DRw13 specificities appended after a decimal point. **Analysis:** The DRw13.02-DQw1 haplotype was less frequent among cases of severe malarial anaemia than mild controls ($RR=0.49$ (0.33-0.72), $\chi^2_{11}=14.0$, $P=0.00018$, $P_{corrected}=0.0023$) but the reduction in frequency among the cerebral malaria group was not significant ($\chi^2_{11}=2.8$). The DR7-DQw2 haplotype was more frequent among cases of severe malarial anaemia than among mild controls ($RR=2.66$ (1.32-5.38), $\chi^2_{11}=8.1$, $P=0.0044$, $P_{corrected}=0.057$). Phenotype frequencies were also analysed. For the DRw13.02 haplotype there were two homozygotes and 34 heterozygotes among the severe malarial anaemics, six homozygotes and 100 heterozygotes among the cerebral malarias, and 19 homozygotes and 120 heterozygotes among the mild controls. Thus, homozygotes for this haplotype had greater protection from severe malarial anaemia ($RR=0.22$) than heterozygotes ($RR=0.55$; χ^2 for trend=13.8, 1 d.f. $P=0.0002$) and homozygotes showed reduced risk of cerebral malaria ($RR=0.36$ (0.12-0.95), $\chi^2_{11}=4.3$, $P=0.04$), but heterozygotes did not ($RR=0.95$ (0.69-1.31)). Of the seven cases of severe malaria who were HbS carriers, two (one of three with cerebral malaria and one of four with severe malarial anaemia) were DRw13.02 heterozygotes. Logistic regression analysis by phenotype⁵⁶, which allowed for possible confounding effects of differences in age, sex, ethnic group and area of residence between the groups of subjects, showed that the overall contribution of all 13 HLA class II haplotypes to differences between the severe malarial anaemic and mild control groups was statistically significant ($\chi^2=25.3$, 13 d.f., $P=0.02$). The adjusted relative risks for DRw13.02-DQw1 and DRw7-DQw2 became, respectively, 0.45 (0.27-0.72) ($\chi^2_{11}=10.6$, $P=0.001$) and 2.42 (1.17-4.99) ($\chi^2_{11}=5.7$, $P=0.02$). The difference in HLA class II haplotype distribution between children with cerebral malaria and mild controls was close to significance ($\chi^2_{11}=22.5$, 14 d.f., $P=0.07$) when the effect of homozygosity for DRw13.02-DQw1 was included. The percentage frequencies of individuals homozygous for any *TaqI* DRB-DQB RFLP pattern in each group were: severe malarial anaemia, 6.9%; cerebral malaria, 14.5%; mild controls, 12.5%; mild malaria, 13.5%; severe controls, 17.0%; healthy adults, 13.2%. The homozygosity rate among the severe malarial anaemia cases was significantly lower than in the mild controls ($\chi^2_{11}=4.4$, $P=0.036$) and severe controls ($\chi^2_{11}=11.0$, $P=0.001$). RFLP analysis using *TaqI* digestion and DRB- and DQB-specific probes was used as the primary HLA class II typing method because, unlike serology, it requires only small amounts of blood and allows DRw13 subtypes to be distinguished and, unlike PCR-oligonucleotide typing methods, it should not miss new alleles in previously unstudied populations. Only the most frequent 13 RFLP-defined haplotypes (all those >2% frequency, accounting for about 85% of all haplotypes) were compared as the others were all too rare to allow a significant association to be detected unless the associated relative risk turned out to be improbably large ($RR > 5$) or small ($RR < 0.2$). The DRB and DQB band pattern labels in Roman numerals are as reported previously^{46,74,75}. (XXI is identical to VI except that the 4.1-kilobase (kb) band is smaller, 4.0 kb; XX is identical to XV except that the 11-kb band is 11.5 kb; XVII differs from XIV only in that the intense 4.0-kb band is 4.1 kb.) The usual complete associations of RFLP patterns with DR types and subtypes, defined by oligonucleotides⁷⁶, were confirmed for this population, except that the common XXI-V haplotype was associated with both DRB1*1304 (78% of alleles) and DRB1*1102 (manuscript submitted). The association of the X RFLP pattern with the DRB1*1302 allele was confirmed for all 48 haplotypes tested using allele-group specific amplification with the primers 5'-TCCTGGACAGATACTCC-3' and 5'-GCACGTGAAGCTCTCAC-3' followed by hybridization with the oligonucleotide probe 5'-ACATCTGGAAGACGAGCG-3'; these primers amplify DRB1*1302 but not DRB1*1301 and a positive signal on subsequent hybridization is specific for DRB1*1302.

Although mechanisms other than pathogen-driven selection might also contribute to the maintenance of MHC polymorphism, this study provides direct evidence that natural selection by infectious diseases exerts an important and continuing evolutionary pressure on HLA genes. □

Received 31 May; accepted 16 July 1991.

1. Dupont, B. (ed.) *Immunobiology of HLA* (Springer, New York, 1989).
2. Klein, J. *Hum. Immunol.* **19**, 155–162 (1987).
3. Serjeantson, S. W. *Transplantation Proc.* **21**, 598–601 (1989).
4. Lawlor, D. A., Ward, F. E., Ennis, P. D., Jackson, A. P. & Parham, P. *Nature* **335**, 268–271 (1988).
5. Figueroa, F., Gunther, E. & Klein, J. *Nature* **335**, 265–267 (1988).
6. Takahata, N. & Nei, M. *Genetics* **124**, 967–978 (1990).
7. Hedrick, P. W. & Thomson, G. *Genetics* **104**, 449–456 (1983).
8. Kiltz, W., Thomson, G. & Baur, M. P. *Am. J. hum. Genet.* **39**, 340–349 (1986).
9. Bjorkman, P. J. *et al.* *Nature* **329**, 512–518 (1987).
10. Parham, P. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 4005–4009 (1988).
11. Hughes, A. L. & Nei, M. *Nature* **335**, 167–170 (1988).
12. Doherty, P. C. & Zinkernagel, R. M. *Lancet* **i**, 1406–1409 (1975).
13. Tiwari, J. L. & Terasaki, P. J. *HLA and Disease Associations* (Springer, New York, 1985).
14. Yamazaki, K. *et al.* *J. exp. Med.* **144**, 1324–1335 (1976).
15. Clarke, B. & Kirby, D. R. S. *Nature* **211**, 999–1000 (1966).
16. Hedrick, P. W. & Thomson, G. *Genetics* **119**, 205–212 (1988).
17. Bruce-Chwatt, L. J. in *Malaria: Principles and Practice of Malarology* 1–59 (eds Wernsdorfer, W. H. & McGregor, I. A. Churchill-Livingstone, Edinburgh, 1988).
18. Playfair, J. H. L. *Br. med. Bull.* **38**, 153–159 (1982).
19. Schofield, L. *et al.* *Nature* **330**, 664–666 (1987).
20. Weiss, W. R. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **85**, 573–576 (1988).
21. Hoffman, S. L. *et al.* *Science* **244**, 1078–1081 (1989).
22. Romero, P. *et al.* *Nature* **341**, 323–325 (1989).
23. Greenwood, B. M. *Parasitol. Today* **6**, 188 (1990).
24. Greenwood, B. M. *et al.* *Trans. Roy. Soc. trop. Med. Hyg.* **81**, 478–486 (1987).
25. Allison, A. C. *Cold Spring Harb. Symp. quant. Biol.* **29**, 137–149 (1964).
26. Sy, N. E., Long, G. W., Alcantara, A. K., Ranoa, C. P. & Cross, J. H. *Southeast Asian J. trop. Med. Public Health* **15**, 1–6 (1984).
27. Restrepo, M., Rojas, W., Montoya, F., Montoya, A. E. & Dawson, D. V. *Rev. Inst. Med. trop. S. Paulo* **30**, 323–331 (1988).
28. Piazza, A. *et al.* *Ann. hum. Genet.* **49**, 47–63 (1985).
29. Brown, P. J. *Hum. Biol.* **53**, 367–382 (1981).
30. Brewster, D., Kwiatkowski, D. & White, N. J. *Lancet* **336**, 313–319 (1990).
31. Warrell, D. A., Molyneux, M. E. & Beales, P. F. *Trans. Roy. Soc. trop. Med. Hyg.* **84**, suppl. 2, 1–65 (1990).
32. MacPherson, G. G. *et al.* *Am. J. Path.* **119**, 385–401 (1988).
33. Bouharoun-Tayoun, H., Attanath, P., Sabchareon, A., Chongsuphajaisiddhi, T. & Drulhe, P. *J. exp. Med.* **172**, 1633–1641 (1990).
34. Allsopp, C. E. M. *et al.* *Hum. Immun.* **30**, 105–109 (1991).
35. Hayashi, H. *et al.* *Immunogenetics* **32**, 195–199 (1990).
36. Okoye, R. C. *et al.* *Tissue Antigens* **25**, 142–155 (1985).
37. Festenstein, H. *et al.* in *Histocompatibility Testing* (eds Dausset, J. & Colombani, J.) 175–178 (Munksgaard, Copenhagen, 1972).

38. Dutoit, E. D., Emmanuel, J. C., West, G., Taljaard, D. G. & Oudshoorn, M. *Tissue Antigens* **36**, 122–124 (1990).
39. du Toit, E. D., MacGregor, K. J., Taljaard, D. G. & Oudshoorn, M. *Tissue Antigens* **31**, 109–125 (1987).
40. Aizawa, M. *HLA in Asia-Oceania* (Hokkaido University Press, Sapporo, 1986).
41. Baur, M. P. *et al.* in *Histocompatibility Testing* (eds Albert, E. D., Baur, M. P. & Mayr, W. R.) 333–341 (Springer, Berlin, 1984).
42. Serjeantson, S. W. in *The Colonization of the Pacific: a Genetic Trail* (eds Hill, A. V. S. & Serjeantson, S. W.) 120–173 (Oxford University Press, Oxford 1989).
43. Cohen, S., McGregor, I. A. & Carrington, S. *Nature* **192**, 733–737 (1961).
44. Bodmer, J. *et al.* *Tissue Antigens* **37**, 97–104 (1991).
45. Lee, K. W., Johnson, A. H. & Hurley, C. K. *J. Immun.* **145**, 3119–3125 (1990).
46. Carlsson, B., Wallin, J., Bohme, J. & Moller, E. *Hum. Immun.* **20**, 95–113 (1987).
47. Bidwell, J. *Immun. Today* **9**, 18–23 (1988).
48. Lee, K. W., Hurley, C. K., Hartzman, R. & Johnson, A. H. *Hum. Immun.* **29**, 202–219 (1990).
49. Fernandez-Vina, M., Shumway, W. & Stastny, P. *Hum. Immun.* **28**, 51–64 (1990).
50. Gorski, J. *et al.* *J. exp. Med.* **170**, 1027–1032 (1989).
51. Schreuder, G. M. T. *et al.* in *Immunobiology of HLA* (ed. Dupont, B.) 252–255 (Springer, New York, 1989).
52. Glynnsten, U. B., Lashkari, D. & Ehrlich, H. A. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1835–1839 (1990).
53. Scharf, S. J., Friedmann, A., Steinmann, L., Brautbar, C. & Ehrlich, H. A. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6215–6219 (1989).
54. Ronningen, K. S. *et al.* *Hum. Immun.* **29**, 275–281 (1990).
55. Lee, T. D., Lee, G. & Zhao, T. M. *Tissue Antigens* **35**, 64–70 (1990).
56. Farewell, V. T. & Dahlberg, S. *Biometrics* **40**, 547–560 (1984).
57. Riley, E. M., Olerup, O. & Troye-Blomberg, M. *Parasitol. Today* **7**, 5–11 (1991).
58. de Vries, R. R., Khan, P. M., Bernini, L. F., van Loghem, E. & van Rood, J. J. *J. Immunogenet.* **6**, 271–287 (1979).
59. Briles, W. E., Stone, H. A. & Cole, R. K. *Science* **195**, 193–195 (1977).
60. Hill, A. V. S. in *Molecular Evolution of the Major Histocompatibility Complex* (eds Klein, J. & Klein, D.) (Springer, Berlin in the press).
61. Haldane, J. B. S. & Jayakar, S. D. *J. Genet.* **58**, 237–242 (1963).
62. Takahata, N. *Genetics* **98**, 427–440 (1981).
63. Gillespie, J. H. *Theor. Populat. Biol.* **27**, 222–237 (1985).
64. Snell, G. D. *Folia Biol.* **14**, 335–358 (1968).
65. Bodmer, W. F. *Nature* **237**, 139–145 (1972).
66. Miettinen, O. S. *Am. J. Epidem.* **99**, 325–332 (1974).
67. Bengtsson, B. O. & Thomson, G. *Tissue Antigens* **18**, 356–363 (1981).
68. Greenwood, B. M., Marsh, K. & Snow, R. W. *Parasitol. Today* (in the press).
69. Molyneux, M., Taylor, T. E., Wirima, J. J. & Borgstein, A. Q. *J. Med.* **72**, 441–459 (1989).
70. Kwiatkowski, D. *et al.* *Lancet* **336**, 1201–1204 (1990).
71. Kwiatkowski, D. & Nowak, M. *Proc. natn. Acad. Sci. U.S.A.* **88**, 5111–5113 (1991).
72. Colombo, B. & Felicetti, L. *J. med. Genet.* **22**, 291–292 (1985).
73. Hill, A. V. S. *et al.* *Lancet* **337**, 640–642 (1991).
74. Martell, R. W., Oudshoorn, M., May, R. M. & du Toit, E. D. *Hum. Immun.* **26**, 237–244 (1989).
75. Martell, R. W., Oudshoorn, M., Arendse, B. & du Toit, E. D. *Hum. Immun.* **28**, 32–38 (1990).
76. Wordsworth, B. P., Allsopp, C. E. M., Young, R. P. & Bell, J. I. *Immunogenetics* **32**, 413–418 (1980).

ACKNOWLEDGEMENTS. We thank the children and their parents and guardians who made this study possible; A. Hughes, V. Sarno, J. Wells, N. J. White, D. Waller, J. Crawley, S. Krishna, C. Craddock, U. Hughes, L. Bayo, L. Manneh, E. G. Sarr and T. Corrah for help with sample collection; M. Hassan-King, I. Sambou, H. Sillah and F. Sisay for assistance in Fajara; C. Taylor and M. Bunce for assistance with serological typing; A. Ting and P. J. Morris for antisera and facilities; S. Abdalla, M. Shepherd and S. N. R. Yates for haemoglobin typing; O. Olerup and C. K. Hurley for discussions and communicating their unpublished data; and J. I. Bell, T. E. A. Peto and D. J. Weatherall for encouragement and advice. A.V.S.H. is a Wellcome Trust Senior Clinical Fellow.

LETTERS TO NATURE

An atomic switch realized with the scanning tunnelling microscope

D. M. Eigler, C. P. Lutz & W. E. Rudge

IBM Research Division, Almaden Research Center, 650 Harry Rd, San Jose, California 95120, USA

THE scanning tunnelling microscope¹ (STM) has been employed in recent years in attempts to develop atomic-scale electronic devices, both by examining device-like characteristics in pre-existing structures^{2,3} and by creating new structures by the precise manipulation of atoms and molecules with the STM tip^{4–6}. Here we report the operation of a bistable switch that derives its function from the motion of a single atom. A xenon atom is moved reversibly between stable positions on each of two stationary conducting 'leads', corresponding to the STM tip and a nickel surface. The state of the switch is set (that is, the xenon atom is moved to the desired location) by the application of a voltage pulse of the appropriate sign across the leads. The state of the switch is identified by measuring the conductance across the leads. This switch is a prototype of a new class of potentially very small electronic devices which we will call atom switches.

The idea that an atom could be transferred from the tip of an STM to the surface by applying a sufficiently high voltage

across the tunnel junction is due to Becker *et al.*⁷, who suggested that the atomic-scale perturbation left on a germanium surface was a germanium atom that had been transferred from the tip to the surface under high bias (tip negative) conditions. Subsequent work used static voltages or voltage pulses to make changes to the tip or to the surface being examined^{8–10}. None of this work demonstrated that an atom can be transferred reversibly between a sample and tip, but Lyo and Avouris¹¹ have recently been able to transfer silicon reversibly between a silicon surface and a tungsten STM tip.

As the conductance of a tunnel junction depends exponentially on the spacing between the electrodes (because the conduction-electron wavefunctions decay exponentially in the tunnel barrier), even slight rearrangements of the atoms in the current-carrying region of the tunnel junction will lead to easily measured changes in the conductance. If controlled reversible motion of an atom between the electrodes of a tunnel junction could be achieved solely by the application of an electrical signal, an atomic-scale bistable electronic switch could be developed.

The atom switches that we will describe were constructed and operated using an STM contained in an ultra-high-vacuum system and cooled to 4 K. The switches were constructed using the (110) surface of a single-crystal nickel sample and a tip made from polycrystalline tungsten wire. The sample was prepared by sputtering with argon ions and annealing to 820 K.

After cooling to 4 K the sample was dosed with a small fraction of a monolayer of xenon atoms. As in most STM studies at present, the chemical identity and structure of the outermost atoms of the tip are not known.

To build our first switch, we used the STM to slide a xenon atom⁴ to a kink site along a monatomic step. We chose this site because we believed that it would enhance the probability that the xenon atom would bind to the same site each time it was transferred to the surface. We then turned off the feedback loop that was used to maintain constant tunnel current, and left the tip at a fixed position above and 5 Å to the side of the xenon atom. The height of the tip was then adjusted to yield a tunnel-junction resistance of 1.5 MΩ with the tip biased at -0.02 V with respect to the sample. In this position, the tip is 3.8 Å above the height of the upper nickel terrace, or 5.0 Å above the lower terrace. It is important to note that the imaging and atomic manipulation modes of the STM, although critical in constructing the switch, are of no importance whatsoever in its operation. During operation of the switch the STM is completely static; its only role is to provide the desired fixed configuration of electrodes.

To identify the conductance state of this switch, we applied -0.02 V to the tip and measured the tunnel current. To toggle the switch to the high-conductance state, in which the xenon atom is bound to the tip, we applied a pulse of +0.8 V to the tip for 64 ms. To toggle to the low-conductance state, in which the xenon atom is bound to the surface, we applied a pulse of -0.8 V for 64 ms (we have also operated a switch with 20 ns pulses). Figure 1 shows the time dependence of the current through this switch during operation.

We verified that the action of the switch was due to the transfer of the xenon atom between the surface and the tip as follows. Starting with the switch in the low-conductance state, we withdrew the tip from the surface until we achieved a resistance of $4 \times 10^7 \Omega$ and then began operating the STM in the constant-current imaging mode. An image of the surface at this time (Fig. 2a) shows the xenon atom bound to the surface. We then returned the tip to its original position near the xenon atom, opened the feedback loop and lowered the tip as before, thus recreating the atom switch conditions. We toggled the switch to the high-conductance state and once again changed to the imaging mode. The image (Fig. 2b) not only reveals the absence of the xenon atom but also shows enhanced resolution of the atomic structure of the nickel surface, indicating a change in the state of the tip, which we believe is due to the presence of the xenon atom on the tip. We then placed the tip (which was then carrying the xenon atom) back over the atom-switch site and toggled the switch to the low-conductance state. Finally we withdrew the tip, and once again imaged the surface (Fig. 2c) to find the xenon atom bound once again to the surface.

We have operated atom switches with xenon atoms located at kink sites, on the bare nickel (110) terrace and on top of single nickel adatoms on the nickel (110) surface. The conductance ratios range from near unity to 7.

We have studied the delay between the onset of the voltage pulse and the change in junction conductance owing to the motion of a xenon atom from a bare nickel (110) terrace to a particular STM tip. For a fixed tip height and pulse voltage, the distribution of delays is a decaying exponential, which indicates a fixed probability of transferring per unit time; that is, there is a characteristic transfer rate. Figure 3 shows that the transfer rate for this particular 906-kΩ junction has a power-law dependence on the pulse voltage, and consequently on the tunnel current, I , according to $I^{4.9 \pm 0.2}$. We were unable to measure the dependence of the transfer rate on the tip-sample separation because at smaller separations (with a junction resistance $R < 700 \text{ k}\Omega$) the xenon atom moved spontaneously to the tip without the need to apply a positive voltage pulse. At larger separations ($R > 1.5 \text{ M}\Omega$) the xenon atom tended to hop among several nearby sites on the nickel surface before transferring to the tip,

and sometimes escaped from the junction region entirely.

We now turn to the question of what physical mechanism causes the motion of the atom. Any candidate mechanism must be an odd function of the applied voltage, have the correct sign, and be consistent with the observed transfer delay statistics.

Lyo and Avouris¹¹ have suggested that ionization followed by field evaporation is the mechanism responsible for the reversible transfer of silicon atoms between the tip and silicon surface of a STM. This mechanism does not explain the xenon transfer, as we find that the motion of the xenon atom is always towards the positively biased electrode.

Mamin *et al.*⁹ have argued that negative-ion formation and subsequent field evaporation can explain the transfer of gold atoms from a negatively biased gold tip to a substrate. Haberland *et al.*¹² have recently presented evidence for the existence of a negative xenon ion with a lifetime greater than $1 \times 10^{-4} \text{ s}$. Negative-ion formation is unlikely to be the mechanism, however, because we do not observe the expected threshold electric field (Fig. 3) and because the existence of a ground-state negative xenon ion would require occupation of the xenon 6s resonance. This resonance lies 4–5 V above the Fermi level¹³ and thus is expected to remain largely empty for the range of voltages used here.

Ralls *et al.*¹⁴ have studied the electromigration¹⁵ of impurities in metal nanobridges. They found that the dominant contribution to the electromigration-force term comes from the scattering of electrons from the impurity and that the electromigration process is enhanced by heating of the impurity above the lattice temperature by inelastic electron scattering at the impurity. Indeed, they suggested that the mechanisms and forces responsible for electromigration of impurities in solids might be responsible for the motion of an atom across the tunnel junction of an STM. Electromigration is odd in the applied field and can often result in motion of the impurity in the direction of electron flow. The competition between electron heating and the relaxation of vibrational energy to the lattice may result in the observed

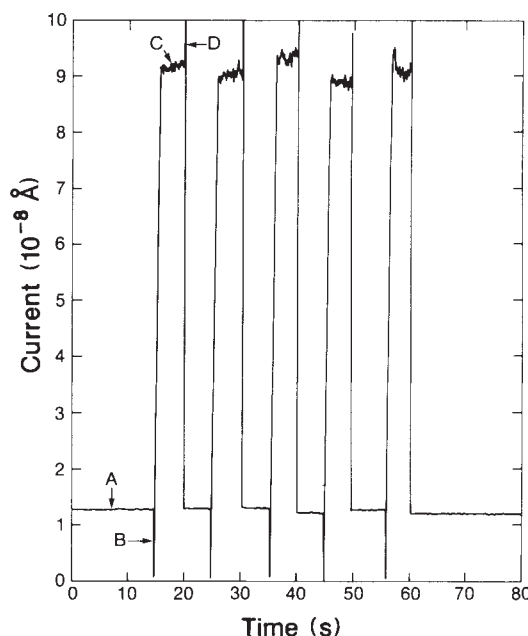


FIG. 1 The time dependence of the current through the switch during operation. We begin (A) in the low-conductance state with the tip biased at -0.02 V and the xenon atom bound to the surface. The transient current spike (B) is due to application of a +0.8-V pulse to the tip for 64 ms, after which the -0.02-V bias to the tip was re-established. This results in the transfer of the xenon to the tip and the establishment of the high-conductance state (C). Applying a -0.8-V pulse for 64 ms yields the transient current spike (D) and re-establishes the low-conductance state by transferring the xenon atom back to the surface. The remainder of the diagram shows the switch repeatedly toggled between the two conductance states.

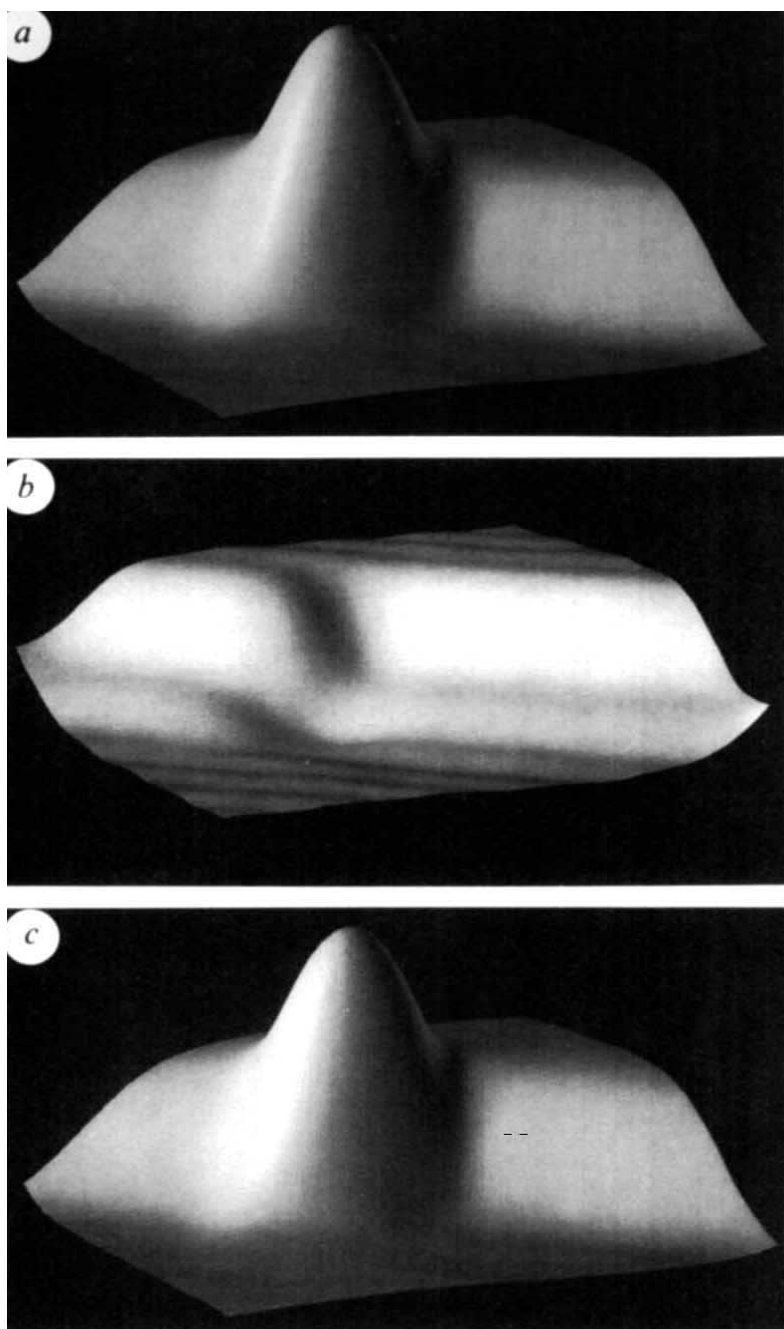
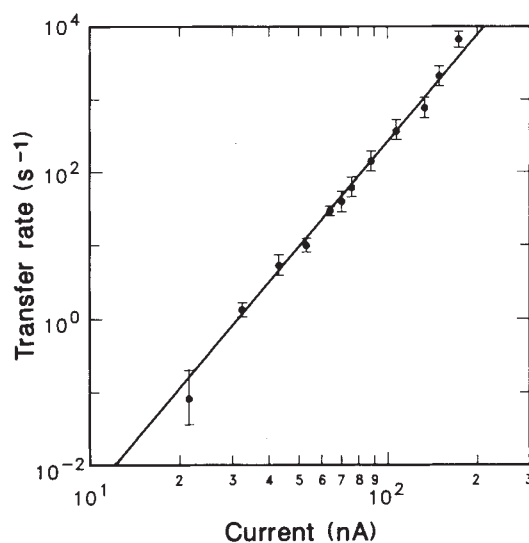


FIG. 2 A sequence of three $25 \times 25 \text{ \AA}$ STM images demonstrating that the operation of the switch is due to the reversible transfer of the xenon atom between the surface and tip. *a*, An image of the surface taken after operation of the switch was interrupted when the switch was in its low-conductance state, showing the xenon atom on the surface. *b*, After toggling the switch to its high-conductance state, in which the xenon atom is on the tip. *c*, After returning the switch to its initial state.

FIG. 3 The measured transfer rate of a xenon atom from the nickel (110) surface to the STM tip as a function of the current, I , during the applied voltage pulse. The transfer rate varies as $I^{4.9 \pm 0.2}$ for this tunnel junction. (For this junction $V/I = 906 \text{ k}\Omega \pm 2\%$ while the xenon atom is on the surface for the full range of voltages used.) A heating-assisted electromigration model is consistent with this behaviour.



power-law dependence on the current, particularly if multiple inelastic scattering events are required. The observed sideways motion of the xenon atom at greater tip-sample separation is consistent with the xenon atom being vibrationally excited. The absence of sideways motion at closer tip-sample separation may be due to the increased van der Waals attraction to the tip as the tip is brought closer to the surface. Heating-assisted electromigration is the only mechanism of which we are aware that is consistent with all the observed phenomena.

Note that the atom switches we have constructed are macroscopic devices in the sense that most of the power dissipation occurs not in the atomic-scale volume of the active element but in a volume comparable to the cube of the inelastic mean free path of electrons in the terminals of the switch. Furthermore, our macroscopic terminals do not show the quantum size effects that atomic-scale leads would exhibit.

It is clear that serious obstacles lie between what we have demonstrated and the realization of a genuinely useful atomic-scale device¹⁶. The prospect of atomic-scale logic and memory devices is nonetheless a little closer. We are intrigued by the idea that atom switches might already exist in the form of single cage-like molecules which derive their switching function from an atom that is trapped in the cage. □

Received 15 July; accepted 23 July 1991.

1. Hansma, P. K. & Tersoff, J. *J. appl. Phys.* **61**, R1-R23 (1987).
2. Lyo, I.-W. & Avouris, P. *Science* **245**, 1369-1371 (1989).
3. Bedrossian, P., Chen, D. M., Mortensen, K. & Golovchenko, J. A. *Nature* **342**, 258-260 (1989).
4. Eigler, D. M. & Schweizer, E. K. *Nature* **344**, 524-526 (1990).
5. *New Scientist* **129**, 20 (23 Feb. 1991).
6. *New Scientist* **26**, 31 (26 Jan. 1991).
7. Becker, R. S., Golovchenko, J. A. & Swartzentruber, B. S. *Nature* **325**, 419-421 (1987).
8. Fuchs, H. & Schimmel, T. *Adv. Mater.* **3**, 112-113 (1991).
9. Mamin, H. J., Guethner, P. H. & Ruger, D. *Phys. Rev. Lett.* **65**, 2418-2421 (1990).
10. Whitman, L. J., Strosio, J. A., Dragoset, R. A. & Celotta, R. J. *Science* **251**, 1206-1210 (1991).
11. Lyo, I.-W. & Avouris, P. *Science* **253**, 173-176 (1991).
12. Haberland, H., Kolar, T. & Reiners, T. *Phys. Rev. Lett.* **63**, 1219-1222 (1989).
13. Eigler, D. M., Weiss, P. S., Schweizer, E. K. & Lang, N. D. *Phys. Rev. Lett.* **66**, 1189-1192 (1991).
14. Ralls, K. S., Ralph, D. C. & Buhrman, R. A. *Phys. Rev. B* **40**, 11561-11570 (1989).
15. Verbruggen, A. H. *IBM J. Res. Dev.* **32**, 93-98 (1988).
16. Landauer, R. *Physica A* **168**, 75-87 (1990).

ACKNOWLEDGEMENTS. We thank D. J. Auerbach for fostering an environment which stimulated this work. We are also grateful to H. Otto, R. B. Martin and J. Downey for their questions.

Electronic states and phases of K_xC_{60} from photoemission and X-ray absorption spectroscopy

C. T. Chen*, L. H. Tjeng*, P. Rudolf*, G. Meigs*, J. E. Rowe*, J. Chen†, J. P. McCauley Jr†, A. B. Smith III†, A. R. McGhie†, W. J. Romanow† & E. W. Plummer†

* AT&T Bell Laboratories, Murray Hill, New Jersey 07974, USA

† Laboratory for Research on the Structure of Matter, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

HIGH-resolution photoemission and soft X-ray absorption spectroscopies have provided valuable information on the electronic structure near the Fermi energy in the superconducting copper oxide compounds¹⁻⁴, helping to constrain the possible mechanisms of superconductivity. Here we describe the application of these techniques to K_xC_{60} , found recently to be superconducting below 19.3 K for $x \approx 3$ (refs 5-7). The photoemission and absorption spectra as a function of x can be fitted by a linear combination of data from just three phases, C_{60} , K_3C_{60} and K_6C_{60} , indicating that there is phase separation in our samples. The photoemission spectra clearly show a well defined Fermi edge in the K_3C_{60} phase with a density of states of 5.2×10^{-3} electrons $eV^{-1} \text{ \AA}^{-3}$ and an occupied-band width of 1.2 eV, suggesting that this phase may be

a weakly coupled BCS-like (conventional) superconductor. The C1s absorption spectra show large non-rigid-band shifts between the three phases with half and complete filling, in the K_3C_{60} and K_6C_{60} phases respectively, of the conduction band formed from the lowest unoccupied molecular orbital of C_{60} . These observations clearly demonstrate that the conduction band has C 2p character. The non-rigid-band shift coupled with the anomalous occupied-band width implies that there is significant mixing of the electronic states of K and C_{60} in the superconducting phase.

A mixture of fullerenes was synthesized by contact arc evaporation⁸ of graphite under 300 torr of He. After Soxhlet extraction with toluene, C_{60} was separated from the mixture by column chromatography on neutral alumina with 5% toluene in hexane as eluant. The purity of C_{60} was >99.9% relative to C_{70} impurity. The sample was bulk-dried under vacuum for several hours at 200 °C, then heated in flowing N_2 to 325 °C before use. Pure C_{60} (ref. 8) was evaporated from a pyrex ampule in an ultrahigh-vacuum chamber and potassium was obtained from a SAES getter source. Both sources were outgassed for several hours after the system was baked. C_{60} was deposited onto a clean Cu(100) substrate at room temperature until a film of thickness $\geq 200 \text{ \AA}$ was obtained. Potassium was evaporated onto the carbon film which was subsequently annealed at 100 °C for ~1 h. After the sample had returned to room temperature, the photoemission and absorption spectra were recorded. The concentration of potassium was controlled by sequential evaporation and annealing steps until a saturated K_6C_{60} sample was obtained. All spectra were measured using the AT&T Bell Laboratories Dragon high-resolution soft X-ray beam line at the National Synchrotron Line Source^{9,10}, with the experimental resolution set at 120 meV for photoemission spectra, 60 meV for absorption spectra.

Figure 1 shows a set of photoemission spectra of the valence-band region of K_xC_{60} as a function of the K concentration. The bottom curve is for pure C_{60} , which is nearly identical to the spectrum published by Weaver *et al.*¹¹. The second curve labelled $\bar{x} = 0.5$ has basically the same features as the pure C_{60} spectrum, except that the small concentration of K has created a Fermi edge 2.6 eV above the highest occupied molecular orbital (HOMO) peak of C_{60} . As the K concentration is increased by sequential evaporation and annealing, the valence-band spectra can be seen to change. The most marked change is the clear

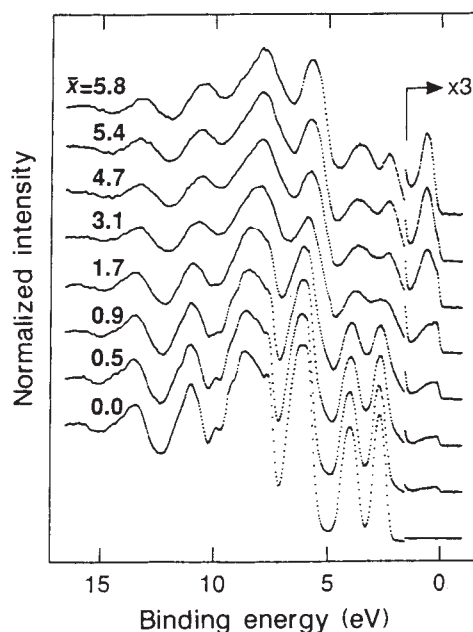


FIG. 1 Valence-band photoemission spectra of K_xC_{60} as a function of x . The Fermi edge region has been enlarged ($\times 3$) with the background subtracted. Energy $h\nu = 110 \text{ eV}$.

development of a Fermi edge; that is, the sample becomes metallic. The region near the Fermi energy (ϵ_F) shows that the sample starts as an insulator, becomes a metal with a maximum density of states (DOS) at ϵ_F for $x=3$, and is nonmetallic at $x=6$. This is consistent with theoretical calculations for both C_{60} (refs 11, 12) and K_6C_{60} (ref. 13), as well as with the conductivity measurements on K_xC_{60} (ref. 14).

Figure 2 shows the soft X-ray absorption spectra for the same sequence of K concentrations depicted in the valence-band spectra of Fig. 1. The lowest-energy peak in the spectrum for C_{60} is the transition from the C 1s to the lowest unoccupied molecular orbital (LUMO). There are significant changes in the soft X-ray absorption spectra as the K concentration increases; the intensity of the excitation to the LUMO of C_{60} decreases while the intensity of the K $2p \rightarrow 3d$ absorption (near $h\nu = 297$ eV) increases.

Two previous reports of the photoemission spectra of K_xC_{60} have been published^{15,16}, but neither of these investigations shows the clear development of the Fermi level intensity as seen in Fig. 1. No soft X-ray absorption spectra have been reported for K_xC_{60} .

The K concentration was determined in three ways: using the intensity of the peak near ϵ_F in Fig. 1, the intensity of the K absorption in Fig. 2 and a fitting procedure applied to the absorption spectra, described later. The average concentration $\bar{x}$ is shown in the figures. The sampling depth in these three measurements is somewhat different, producing slight variations (of the order of $\pm 10\%$) in the measured concentration.

The data presented in Fig. 1 for the valence-band structure near ϵ_F indicate that there are only three phases: C_{60} with no structure within ~ 2 eV of ϵ_F , K_3C_{60} with a conduction band crossing ϵ_F and K_6C_{60} with a filled band ~ 0.6 eV below ϵ_F . Figure 3 shows this region of the spectra for these three phases. The curve for K_3C_{60} was generated from the $\bar{x}=3.1$ curve by removing a small contribution from the saturated spectrum. The DOS at ϵ_F can be directly measured from the spectral intensity at ϵ_F after the instrumental broadening has been removed. The conversion of the intensity scale to states per electron volt per C_{60} molecule is accomplished by using the HOMO of the K_6C_{60} as a reference. This band is derived from the LUMO of C_{60} which contains three states (six electrons). An alternative normalization using the area of the conduction band itself gives almost identical results. Our measured DOS at ϵ_F is 1.9 ± 0.1 states per eV per K_3C_{60} , equivalent to 5.2×10^{-3} electrons $eV^{-1} \text{ \AA}^{-3}$. To extract the occupied-band width of the conduction

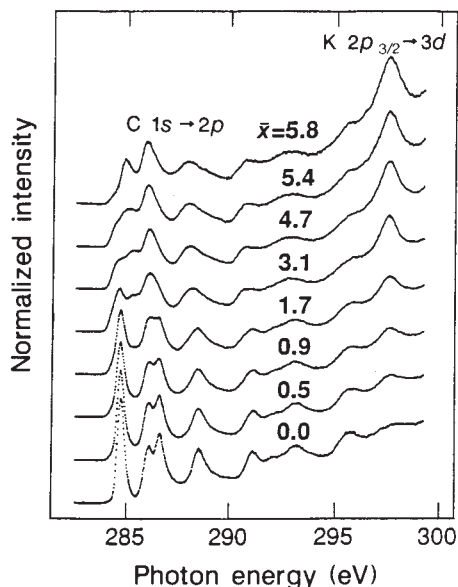


FIG. 2 Soft X-ray absorption spectra of K_xC_{60} as a function of x .

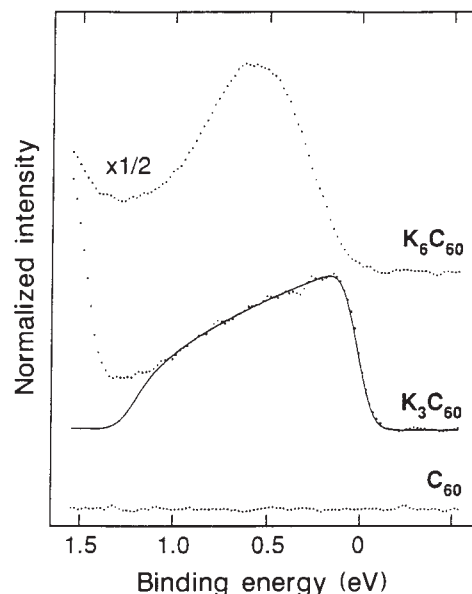


FIG. 3 Photoemission spectra near the Fermi edge for three phases of K_xC_{60} ($h\nu=110$ eV).

band of K_3C_{60} , the data are fitted with a free-electron-like (FE) band, $N(\epsilon) = N(\epsilon_F) [(\epsilon - \epsilon_0)/(\epsilon_F - \epsilon_0)]^{1/2}$, convoluted with a gaussian instrumental response function (solid line). The occupied-band width is found to be 1.2 ± 0.05 eV. As this band is derived from the C $2p$ atomic orbitals, a tight-binding-like (TB) band has also been used to fit the data. The standard deviation of the TB fit (not shown) is two times larger than that of the FE fit, but gives a similar occupied-band width. This occupied-band width in K_3C_{60} is almost the same as the full-band width of the HOMO in K_6C_{60} . In weak-coupling BCS theory, the transition temperature is related to the DOS by the relationship $kT_c = 1.13E \exp[-1/(N(\epsilon_F)V_0)]$, where E is the energy of the pairing-mediation excitation, V_0 is the electron-excitation coupling strength and $\lambda = NV_0$ is the coupling constant. Cheng and Klein¹⁷ have calculated the phonon modes for both the face-centred cubic and body-centred cubic phases of K_3C_{60} , observing that the low-energy translational and librational motions of the C_{60} molecules are strongly coupled to the K motion. If we use the energy of 30 cm^{-1} found for the face-centred cubic phase, we obtain a value of $\lambda = 1.06$, which is large for weak coupling. This value of λ is then due to the low frequency of the coupled librational-translational modes and should be considered as an upper limit. Even for this λ we have $V_0 = 0.56$ eV, corresponding to a deformation potential $\partial V/\partial u = 0.27 \text{ eV \AA}^{-1}$. This is a relatively modest matrix element for electron-phonon scattering, consistent with K_3C_{60} being a BCS superconductor.

The original X-ray diffraction studies of K_xC_{60} powders indicated that for $0 < x < 6$ the sample was almost completely phase-separated into C_{60} and K_6C_{60} (ref. 18), consistent with the first report of superconductivity in K_3C_{60} , where it was claimed that only 1% of the sample contributed to the superconducting phase⁵. Recently, Holczer *et al.*⁷ showed that there was a single superconducting phase K_3C_{60} , which is now believed to be face-centred cubic with K incorporated into all of the octahedral and tetrahedral interstices¹⁹. All of the spectral features near ϵ_F in Fig. 1 can be fitted with the three pure-phase spectra shown in Fig. 3. It is slightly more difficult to fit the absorption spectra in a similar fashion, because the three phases have markedly different structures. The absorption spectrum for pure K_3C_{60} was constructed from the difference between the spectrum for $\bar{x}=0.9$ and the pure C_{60} spectrum and from the difference between the last two curves in the doping sequence. The assumption was that if there is phase separation, as indicated by the

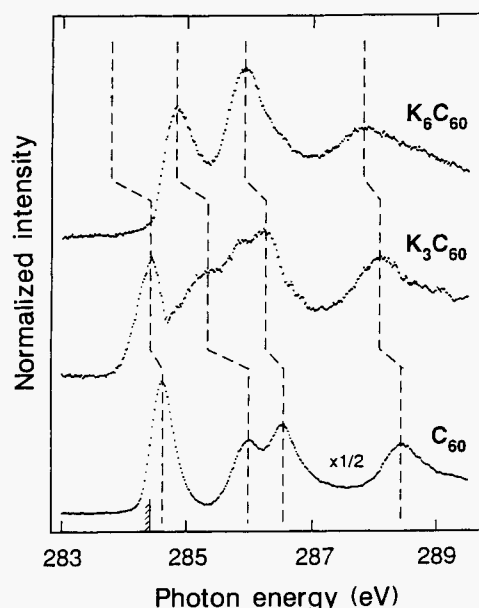


FIG. 4 C 1s absorption spectra for the three phases of K_xC_{60} . The shaded vertical line at the bottom indicates the C 1s excitation threshold for the superconducting phase.

valence photoemission curves, then this procedure should give the same spectrum for K_3C_{60} in the limits of low and high K concentration. This assumption is justified by internal consistency tests of the absorption data, where all spectra can be fitted by combinations of the three pure-phase spectra, C_{60} , K_3C_{60} , and K_6C_{60} , shown in Fig. 4.

The width of the peak in the K_3C_{60} absorption spectrum at the threshold is due almost entirely to the width of the carbon core level. This sharp peak indicates that there is an edge singularity effect in the excitation spectrum from K_3C_{60} , that is, a metallic exciton. Energy bands of similar character are found to shift toward the excitation threshold in the three phases. The area of the threshold peak in the K_3C_{60} spectrum is exactly half the area of the LUMO in C_{60} . The important observation is that there is not a rigid-band shift of the C_{60} orbitals in these three phases. These observations clearly demonstrate the $C 2p$ character of the energy band near ϵ_F observed in photoemission. Coupled with the anomalous occupied-band width in K_3C_{60} , these results imply a considerable hybridization between the electronic states of K and C_{60} in the superconducting phase. Presumably, this hybridization is responsible for the electron-phonon coupling, which together with the relatively high value of the DOS at ϵ_F suggests that K_3C_{60} may be a weak-coupling BCS-type superconductor. □

Received 26 June; accepted 26 July 1991.

1. Fujiomori, A., Takayama-Muromachi, E., Uchida, Y. & Okai, B. *Phys. Rev. B* **35**, 8814 (1987).
2. van der Marel, D., van Elp, J., Sawatzky, G. A. & Heitmann, D. *Phys. Rev. B* **37**, 5136 (1988).
3. Shen, Z.-X. *et al. Phys. Rev. B* **36**, 8414 (1987).
4. Chen, C. T. *et al. Phys. Rev. Lett.* **66**, 104 (1991).
5. Hebard, A. F. *et al. Nature* **350**, 600–601 (1991).
6. Rosseinsky, M. J. *et al. Phys. Rev. Lett.* **66**, 2830–2832 (1991).
7. Holczer, K. *et al. Science* **252**, 1154–1157 (1991).
8. Krätschmer, W., Lamb, L. D., Fostiropoulos, K. & Huffman, D. R. *Nature* **347**, 354–358 (1990).
9. Chen, C. T. *Nucl. Instr. Methods A* **256**, 595 (1987).
10. Chen, C. T. & Settle, F. *Rev. Sci. Instrum.* **60**, 1616–1621 (1989).
11. Weaver, J. H. *et al. Phys. Rev. Lett.* **66**, 1741–1744 (1991).
12. Martins, J. L., Troullier, N. & Weaver, J. H. *Chem. Phys. Lett.* **180**, 457–460 (1991).
13. Erwin, S. C. & Pederson, M. R. *Phys. Rev. Lett.* (submitted).
14. Haddon, R. C. *et al. Nature* **350**, 320–322 (1991).
15. Benning, P. J., Martins, J. L., Weaver, J. H., Chibante, L. P. F. & Smalley, R. E. *Science* **252**, 1417–1419 (1991).
16. Wertheim, G. K. *et al. Science* **252**, 1419–1521 (1991).
17. Cheng, A. & Klein, M. L. *Phys. Rev. Lett.* (submitted).
18. Zhou, O. *et al. Nature* **351**, 462–464 (1991).
19. Stephens, P. W. *et al. Nature* **351**, 632–634 (1991).

ACKNOWLEDGEMENTS. We thank E. J. Mele, G. K. Wertheim and J. Zaanen for discussions. This research was partially funded by NSF. The National Synchrotron Light Source is supported by the DOE.

Magnetic-field penetration depth in K_3C_{60} measured by muon spin relaxation

Y. J. Uemura*, A. Keren*, L. P. Le*, G. M. Luke*, B. J. Sternlieb*, W. D. Wu*, J. H. Brewer†, R. L. Whetten‡, S. M. Huang‡, Sophia Lin‡, R. B. Kaner‡, F. Diederich‡, S. Donovan§, G. Grüner§ & K. Holczer§

* Department of Physics, Columbia University, New York, New York 10027, USA

† TRIUMF and Department of Physics, University of British Columbia, Vancouver, British Columbia, V6T 2A3, Canada

‡ Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90024, USA

§ Department of Physics, University of California, Los Angeles, California 90024, USA

THE discovery^{1–3} of superconductivity in C_{60} doped with the alkali metals potassium and rubidium has introduced a new family of three-dimensional molecular superconductors⁴. The potassium-doped compound³ K_3C_{60} has a relatively high transition temperature ($T_c = 19.3$ K), a very high upper critical field ($H_{c2}(T \rightarrow 0) \approx 50$ T) and a short superconducting coherence length⁵ ($\xi = 26$ Å), in common with the copper oxide superconductors. Here we report muon-spin-relaxation measurements of the magnetic-field penetration depth λ in K_3C_{60} . The temperature dependence of λ and of the muon spin relaxation rate indicate that the superconducting energy gap is isotropic, without nodes or zero points. The low-temperature penetration depth $\lambda(T \rightarrow 0)$ is about 4,800 Å, which implies a ratio of superconducting carrier density to effective mass to be $n_s/(m^*/m_e) = 1.2 \times 10^{20} \text{ cm}^{-3}$ if one assumes the 'clean limit'. Combining this result with the value of ξ , we estimate the Fermi temperature $T_F = 470$ K. In the relationship between T_F and T_c , K_3C_{60} conforms to the trend exhibited by 'exotic' superconductors^{6,7} such as the Chevrel phase compounds, the copper oxides and the organic BEDT systems.

The muon-spin-relaxation (μ SR) technique has been extensively applied in the study of the penetration depth in various type-II superconductors^{8–10}. In transverse-field μ SR (TF- μ SR) measurements of λ , we apply an external magnetic field H_{ext} ($H_{c1} \ll H_{ext} \ll H_{c2}$), and observe the spin precession of positive muons implanted in the specimen. In the superconducting state, the field H_{ext} forms a lattice of flux vortices in the specimen, resulting in a local magnetic field B having a distribution with a width ΔB proportional to λ^{-2} . In the spectra of muon spin precession, the oscillation amplitudes are damped because of the inhomogeneity of the local field B . This damping is usually described by a gaussian envelope $G_x(t) = \exp(-\frac{1}{2}\sigma^2 t^2)$ which defines the muon spin relaxation rate $\sigma \propto \Delta B \propto \lambda^{-2}$.

The specimen of K_3C_{60} was prepared at UCLA following the procedures described in ref. 3. A polycrystalline powder material weighing 135 mg was pressed into a sintered pellet ~ 1 cm in diameter and 1 mm thick. Before pressing, the powder sample showed a shielding diamagnetism ranging from 40 to 60% of a bulk niobium reference. The difference from the reference is mostly due to the small packing density of the powder. Pressing and sintering results in an onset of bulk (100%) shielding at a temperature ~ 1 K lower than $T_c = 19.3$ K (ref. 11). X-ray studies on a similar specimen⁴, sensitive to microscopic-scale inhomogeneity, set an upper limit of 15% volume fraction for presumably nonsuperconducting minority phases. The sintered specimen for μ SR was mounted in a ^4He gas flow cryostat with its face normal to the beam direction along which the external field was applied. During the whole procedure of sample preparation and μ SR measurements, the specimen was kept under vacuum or in dry He gas, except for an interval of less than 1

minute during its loading into the cryostat. A polarized positive muon beam from the M15 channel of TRIUMF was stopped in the specimen, and the time spectra of muon decay events were recorded with an apparatus described in ref. 10.

We made two sets of measurements with a transverse external field of $H_{\text{ext}} = 1$ kG and 2 kG applied perpendicular to the initial muon-spin polarization. Figure 1 shows the precession signal in the time spectra plotted using a rotating reference frame. We observed a long-lived oscillation in H_{ext} above T_c and evident depolarization below T_c on cooling because of the inhomogeneous field distribution in the vortex state. There is only a minimal long-lived precessing component remaining after $t \approx 6 \mu\text{s}$ at $T = 3.3$ K. Generally, signals from superconducting and nonsuperconducting regions in the specimen combine additively to give the total μSR signal, with initial precession amplitudes proportional to their volume fractions. The data at $T = 3.3$ K in Fig. 1 indicate that the volume of nonsuperconducting region contributing to the background long-lived signal is less than 10% of the total volume. The present data can be fitted reasonably well with a gaussian curve for $G_x(t)$. After correcting for a small depolarization due to nuclear dipolar fields $\sigma_{\text{nd}} = 0.08 \mu\text{s}^{-1}$ observed above T_c , we obtained the flux-lattice-induced relaxation rate σ from the observed relaxation rate σ_{obs} as $\sigma^2 = \sigma_{\text{obs}}^2 - \sigma_{\text{nd}}^2$.

The temperature dependence of the relaxation rate σ is shown in Fig. 2a. In the field-cooling (FC) measurements, σ increases with decreasing temperature below T_c , and then saturates at low temperatures. There is essentially no dependence of σ on H_{ext} , as expected from theories^{12,13} in which ΔB is due to flux-vortex lattice. Separate zero-field μSR measurements show no effect of static magnetic order, confirming that σ is due purely to superconductivity. The corresponding values of the penetration depth $\lambda(T)$ are shown in Fig. 2b. The low-temperature penetration depth $\lambda(T \rightarrow 0)$ is $\sim 4,800 \text{ \AA}$, with a systematic error due to possible nonsuperconducting volume less than $\pm 200 \text{ \AA}$. Note that the value of λ could vary systematically by up to about $\pm 15\%$ depending on different models for field distributions and functional forms of $G_x(t)$. For the conversion factor α in $\lambda = \alpha \sigma^{-1/2}$ with λ in $\text{\AA}$ and σ in μs^{-1} , Pincus *et al.*¹² (after modifying for the triangular lattice) obtained $\alpha = 2,700$. The Brandt model¹³ gives $\alpha = 3,270$ if we substitute $\sqrt{M_2}$, where

M_2 is the second moment of the field distribution, as σ/γ_μ (γ_μ is the gyromagnetic ratio of μ^+). The 'high-field tail' in the Brandt field distribution, contributing heavily to M_2 , is not well represented when we fit $G_x(t)$ with a gaussian shape. Thus the fitted value of σ for the Brandt distribution corresponds to $\sim 0.7\sqrt{M_2} \times \gamma_\mu$, according to our numerical simulations. The Brandt model then also yields $\alpha \approx 2,700$. We use this conversion factor.

The μSR result for λ is somewhat larger than $\lambda(T \rightarrow 0) \approx 2,400 \text{ \AA}$ estimated from $H_{c1} \approx 130 \text{ G}$ (ref. 5). We note that H_{c1} generally tends to be overestimated because of the effects of flux pinning, yielding underestimates for λ . In contrast, μSR results in field-cooling measurements are not affected by flux pinning. As we discuss later, we found evidence for flux pinning in K_3C_{60} even at temperatures very close to T_c . Another possible source of the difference in the λ values in μSR and H_{c1} studies is the small grain size, typically $\sim 10,000 \text{ \AA}$. The closeness of λ to the grain size could significantly affect the H_{c1} results. The approximate vortex separation in the fields $H_{\text{ext}} \approx 1\text{--}2 \text{ kG}$ used in this μSR study is $\sim 1,000 \text{ \AA}$. About 100 vortices exist in a grain, suggesting that the effect of grain size may be less important in μSR than in H_{c1} measurements.

As $\sigma \propto \lambda^{-2}$ is proportional to the density of superconducting carriers n_s , which screen the external field, thermal pair-breaking excitations across the energy gap reduce n_s and σ , and λ therefore increases at finite temperatures. This feature allows study of the gap symmetry by μSR . Analysing the observed temperature dependence $\sigma(T)$ with a trial formula $[1 - (T/T_c)^\alpha]$, we obtain the best fit for $\alpha = 3.2$ (solid line in Fig. 2a) and $T_c = 18.9 \text{ K}$. This value of α is close to $\alpha = 4$ expected in the two-fluid model¹⁴ for isotropic superconductors. As shown in Fig. 2b, the observed results of $\sigma(T)$ lie between the curves for the two-fluid model and for the weak-coupling BCS model¹⁴ for s-wave supercon-

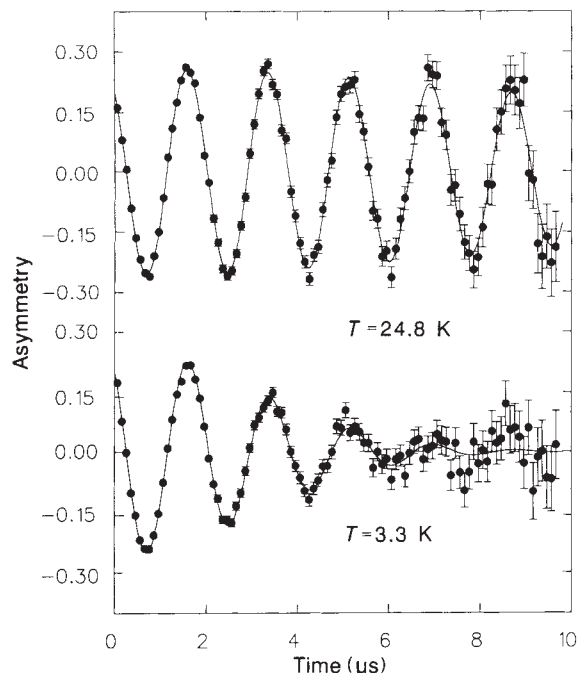


FIG. 1 Muon-spin-precession spectra in K_3C_{60} observed in a transverse external magnetic field $H_{\text{ext}} = 2 \text{ kG}$ after field cooling.

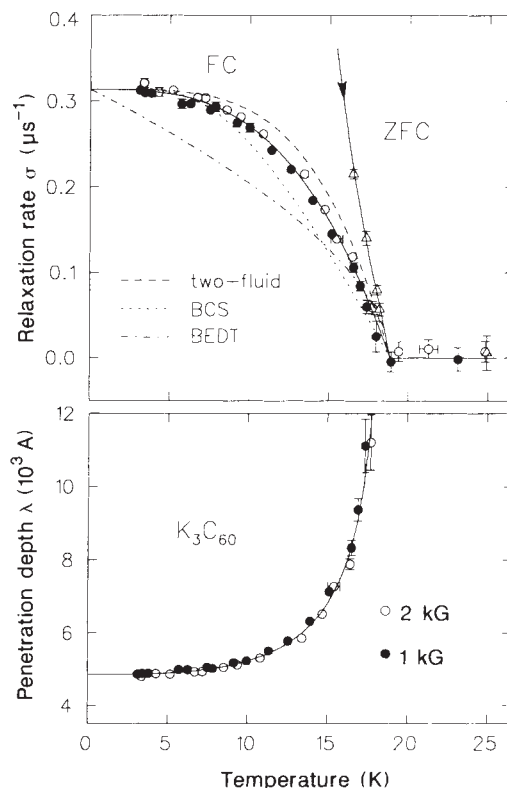


FIG. 2 a, Temperature dependence of the muon-spin-relaxation rate $\sigma(T)$ in K_3C_{60} . The solid line shows a fit to $\sigma(T) = [1 - (T/T_c)^{3.2}]$ with $T_c = 18.9 \text{ K}$. Also shown are curves expected for the two-fluid model, the BCS weak-coupling model and a recent μSR result in the BEDT system after normalizing by T_c . b, The magnetic field penetration depth λ in K_3C_{60} derived from $\sigma \propto \lambda^{-2}$.

ductors. This result indicates that K_3C_{60} has an isotropic energy gap without anomalous zeros: the system is most likely to be an s-wave superconductor.

This feature is common to the high- T_c cuprate, bismuthate (BKBO), Chevrel phase (see refs 6–10), and many other superconductors. For comparison, we show a curve of $\sigma(T)$ observed in $(BEDT-TTF)_2Cu(NCS)_2$ in a recent μ SR study^{10,15} in Fig. 2a. This curve has a linear variation in T at low temperatures, characteristic of anisotropic superconductors with line nodes in the energy gap^{10,15}. The present results for $\sigma(T)$ in K_3C_{60} show a distinctly different temperature dependence with saturation at low temperatures, thus demonstrating isotropic pairing.

In general, the penetration depth λ is given as a function of n_s , m^* , ξ and the mean free path l as

$$\frac{1}{\lambda^2} = \frac{4\pi n_s e^2}{m^* c^2} \times \frac{1}{1 + \xi/l}.$$

For systems close to the clean limit, $\xi/l \rightarrow 0$, the second term essentially becomes unity, and one expects a simple relation $\sigma \propto \lambda^{-2} \propto n_s/m^*$. Estimates of ξ from H_{c2} measurements and l from quantum oscillation and/or a.c.-conductivity studies indicate that the cuprate, organic BEDT, UPT₃ and some other systems lie close to the clean limit with $\xi/l \leq 0.3$. For K_3C_{60} , $\xi = 26 \text{ \AA}$ was obtained⁵ from $H_{c2} \approx 50 \text{ T}$, while no reliable estimates are currently available for l . In view of the extremely short coherence length, here we assume that K_3C_{60} also lies close to the clean limit. With this assumption, we obtain the ground-state value $n_s/(m^*/m_e) \approx 1.2 \times 10^{20} \text{ cm}^{-3}$. In a plot of T_c against $\sigma(T \rightarrow 0)$ comparing μ SR results from different compounds^{6,7}, the present results from K_3C_{60} give a point for this system lying very close to the linear relation found for the cuprates, organic BEDT and some other systems. This implies

that the ratio between T_c and $\sigma \propto n_s/m^*$ is roughly the same for these superconductors.

The Fermi temperature T_F can be derived from $\sigma \propto n_s/m^*$. In three-dimensional systems, σ has to be combined with one other parameter containing either n_s or m^* , or their combination, to calculate $T_F \propto n_s^{2/3}/m^*$. As described in ref. 7, the most convenient way is to use the Sommerfeld constant $\gamma \propto n_s^{1/3} m^*$, to obtain $T_F \propto n_s^{2/3}/m^* \propto \sigma^{3/4} \gamma^{-1/4}$. In K_3C_{60} , a strong contribution from low-energy vibrational modes in the specific heat prevents a reliable determination of γ . Here we use the coherence length $\xi = (\hbar v_F)/(\Delta_G \pi)$, assuming the energy gap $\Delta_G = 1.76 k_B T_c$, which gives the Fermi velocity $v_F = 3.6 \times 10^6 \text{ cm s}^{-1}$. Combining $\sigma \propto n_s/m^*$ and $v_F \propto n_s^{1/3}/m^*$, we obtain $T_F = 470 \text{ K}$ assuming the formula

$$k_B T_F = (\hbar^2/2)(3\pi^2)^{2/3} n_s^{2/3}/m^* \propto \sigma^{1/2} v_F^{1/2}$$

for a noninteracting free-electron gas. The value of T_F thus calculated does not necessarily correspond to the Fermi temperature in detailed band-structure calculations, but does represent a characteristic energy scale for the superconducting carriers in the system. The same combination of parameters also yields $m^* = 11 m_e$ and $n_s = 1.4 \times 10^{21} \text{ cm}^{-3}$, which is close to the carrier density expected for one charge carrier per C_{60} molecule. Thus, K_3C_{60} has a narrow band with a low carrier density.

In Fig. 3 we compare the results from K_3C_{60} with those from other systems⁷ in a plot of T_c against T_F . No corrections are made for ξ/l in calculating T_F : these would shift points towards the right-hand side on the horizontal axis, but would not alter the essential features. K_3C_{60} lies along the linear trend common to the 'exotic' superconductors defined in ref. 7. The ratio $T_c/T_F = 1/25$ in K_3C_{60} is fairly large and close to those in high- T_c cuprate and two-dimensional organic BEDT superconductors, although these systems have different crystal and electronic structures and dimensionalities. Implications of the linear trend have been discussed in ref. 7.

Finally, we report on flux-pinning properties of K_3C_{60} . In zero-field cooling (ZFC) measurements with H_{ext} , flux lines cannot reach an equilibrium configuration if they are pinned in the process of moving from the edge to the inside of the specimen. This broadens the local field and produces a corresponding increase in σ . Therefore, the flux-pinning temperature T_p can be determined by μ SR as the temperature below which the ZFC values of σ deviate from the FC values. As shown in Fig. 2a, values of $\sigma(T)$ measured in ZFC are larger than those in FC even at temperatures very close to T_c indicating $T_p/T_c \geq 0.95$ at $H_{ext} = 2 \text{ kG}$ in K_3C_{60} . Earlier studies^{16,17} suggest a tendency for systems with larger two-dimensional anisotropy, such as the Bi2212 cuprate or $(BEDT-TTF)_2Cu[N(CN)_2]Br$ (ref. 15), to have lower values of T_p/T_c . The present finding in the isotropic K_3C_{60} is consistent with this empirical trend. □

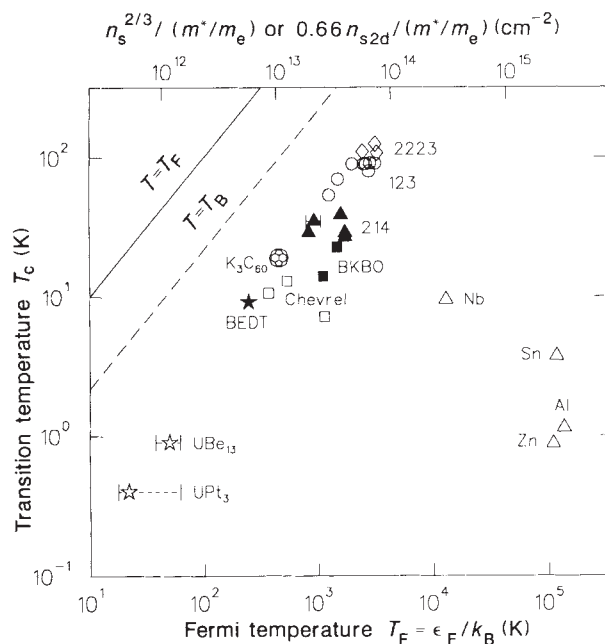


FIG. 3 A plot of T_c against the Fermi temperature T_F derived from our penetration depth measurements in K_3C_{60} (soccer-ball symbol) and various other superconductors (ref. 7 and refs therein). Open diamonds (2223) denote $Bi_2Sr_2(Ca,Pb)_2Cu_3O_{10}$ and cuprate compounds with similar structures; open circles (123) denote $YBa_2Cu_3O_y$ with $y = 6.67-7.0$ and $Y_{0.7}Ca_{0.3}Ba_2Cu_3O_7$; filled triangles (214) denote $La_{2-x}Sr_xCuO_4$ with $x = 0.1-0.21$; filled squares (BKBO) denote $Ba_{1-x}K_xBiO_3$ with $x = 0.4-0.5$; open squares (Chevrel) denote $PbMo_6S_8$ and other nonmagnetic Chevrel-phase compounds. The broken line corresponds to the Bose-Einstein condensation temperature T_B for an ideal three-dimensional boson gas with a boson density $\frac{1}{2}n_s$ and mass $2m^*$. n_s is defined in the text; n_{s2d} is a two-dimensional number density (see ref. 7 for details).

Received 1 July; accepted 23 July 1991.

1. Hebard, A. F. *et al.* *Nature* **350**, 600–601 (1991).
2. Rosseinsky, M. J. *et al.* *Phys. Rev. Lett.* **66**, 2830–2832 (1991).
3. Holczer, K. *et al.* *Science* **252**, 1154–1157 (1991).
4. Stephens, P. W. *et al.* *Nature* **351**, 632–634 (1991).
5. Holczer, K. *et al.* *Phys. Rev. Lett.* **67**, 271–274 (1991).
6. Uemura, Y. J. *et al.* *Phys. Rev. Lett.* **62**, 2317–2320 (1989).
7. Uemura, Y. J. *et al.* *Phys. Rev. Lett.* **66**, 2665–2668 (1991).
8. Keller, H. *IBM J. Res. Dev.* **33**, 314–322 (1989).
9. Uemura, Y. J. *et al.* *Physica C* **162–164**, 857–860 (1989).
10. Uemura, Y. J. *Physica B* **169**, 99–106 (1991).
11. Sparr, G. *et al.* *Science* **252** (in the press).
12. Pincus, P., Gossard, A. C., Jaccarino, V. & Wernick, J. H. *Phys. Lett.* **13**, 31–33 (1964).
13. Brandt, E. H. *Phys. Rev. B* **37**, 2349–2352 (1988).
14. Rickayzen, G. *Theory of Superconductivity* (Wiley, New York, 1965).
15. Le, L. P. *et al.* *High-Temperature Superconductor-M²S III Conf.*, *Physica C* (in the press).
16. Pümpin, B. *et al.* *Z. Phys.* **B72**, 175 (1988).
17. Sternlieb, B. J. *et al.* *Physica C* **162–164**, 679–680 (1989).

ACKNOWLEDGEMENTS. S. J. Anz and F. Ettl produced and separated the molecular carbon samples used here. This work is supported by the David and Lucile Packard Foundation (Y.J.U., R.L.W. and R.B.K.), the NSF, the Office of Naval Research (F.D.) and the Natural Science and Engineering Research Council of Canada.

Observation of speckle by diffraction with coherent X-rays

M. Sutton*, S. G. J. Mochrie†, T. Greytak‡,
S. E. Nagler‡, L. E. Berman§, G. A. Held||
& G. B. Stephenson||

* Department of Physics, Centre for the Physics of Materials,
McGill University, Montréal, Québec, Canada H3A 2T8

† Department of Physics, Massachusetts Institute of Technology,
Cambridge, Massachusetts 02139, USA

‡ Department of Physics, University of Florida, Gainesville,
Florida 32611, USA

§ National Synchrotron Light Source, Brookhaven National Laboratory,
Upton, New York 11973, USA

|| IBM Research Division, Thomas J. Watson Research Center,
Yorktown Heights, New York 10598, USA

REFLECTED light from a coherent light source such as a laser shows a graininess known as speckle. In general, a speckle pattern is produced whenever randomly distributed regions of a material introduce different phase shifts into the scattering of coherent incident light. If the arrangement of the regions evolves with time, the speckle pattern will also change. Observation of the intensity fluctuations at a single point in the speckle pattern provides a direct measure of the time correlation function of the inhomogeneity. This leads to a technique^{1,2}, alternatively called light-beating spectroscopy, dynamic light scattering or intensity fluctuation spectroscopy, which has been widely used with visible light to study processes such as critical fluctuations near phase transitions in fluids and the diffusion of particles in liquids. But it is not possible to study processes involving length scales less than about 200 nm, or those in opaque materials, using visible light. If intensity fluctuation spectroscopy could be carried out using coherent X-rays, however, one could probe the dynamics of processes involving atomic length scales in a wide range of materials. Here we show that by using a high-brilliance X-ray source it should be possible to perform this type of measurement using radiation of wavelength ~ 0.15 nm. Specifically, we have observed a speckle pattern in the diffraction of coherent X-rays from randomly arranged antiphase domains in a single crystal of the binary alloy Cu_3Au .

To obtain a speckle pattern, the incident beam must be sufficiently coherent. With visible light this is typically achieved with a laser source, but it is possible to use an incoherent source, such as a mercury arc lamp² or, as in our case, an X-ray wiggler, provided that it is suitably collimated and monochromatic. The coherent intensity so obtained depends directly on the brilliance of the source. To resolve speckle, the difference in the path lengths of rays scattered by different parts of the sample must not be much greater than the longitudinal coherence length of the incident beam. This is equal to $\lambda^2/2\delta\lambda$, where λ is the wavelength and $\delta\lambda$ is the distribution of wavelengths. In addition, the incident beam size must not be much greater than the transverse coherence length³, given by $\lambda R_s/2d_s$, where R_s is the distance from the source and d_s is the source size. For a typical X-ray wavelength, $\lambda = 0.15$ nm, a silicon (111) monochromator gives $\delta\lambda/\lambda \approx 1.4 \times 10^{-4}$, which results in a longitudinal coherence length of ~ 0.5 μm . For a 0.3-mm-diameter synchrotron source at $R_s = 30$ m, the transverse coherence length is ~ 7 μm . Thus an X-ray wiggler source with a brilliance of 10^{15} photons $\text{s}^{-1} \text{ mrad}^{-2} \text{ mm}^{-2}$ per 0.1% bandwidth, when collimated through a 5 μm pinhole, is expected to deliver a usable coherent beam of $\sim 3 \times 10^5$ photons s^{-1} . As we shall demonstrate, it is possible to obtain sufficient coherent intensity with current wiggler X-ray sources to perform scattering measurements from certain types of samples.

The experiments were performed using the X25 wiggler beamline at the national synchrotron light source (NSLS). Figure 1

shows the scattering geometry. The nominal size of the wiggler source is 0.4 mm in the horizontal and 0.1 mm in the vertical⁴. A wavelength of 0.155 nm was selected by a specially designed double crystal silicon (111) monochromator⁵. A coherent beam of X-rays was obtained by placing either a 2.5- μm or 5 μm nominal diameter pinhole 28 m from the source. The collimated beam was then diffracted by the sample, which was located 40 mm behind the pinhole. A second, analysing pinhole was placed 1 m from the sample and the scattered intensity profile measured by counting the photons that passed through this pinhole as it was scanned in and perpendicular to the scattering plane (the 2θ and α directions, respectively).

To verify that our X-ray source provided the required transverse coherence, we measured the Fraunhofer diffraction pattern of the collimating pinhole by removing the sample and bringing the detector angle, 2θ , to near zero. Figure 2 shows the diffraction patterns in the 2θ direction for the 2.5- and 5- μm pinholes. Diffraction profiles in the α direction are similar. The solid lines are the fits to Fraunhofer diffraction patterns of circular apertures³ convolved with the acceptance of the analysing pinhole. The fitted aperture diameters of 2.9 and 4.5 μm agree well with the pinhole sizes measured by electron microscopy. The observed shape of the central peak and the positions of the subsidiary minima and maxima of the measured patterns agree well with the theoretical curve, indicating that the transverse coherence length of the source is at least 5 μm in both the horizontal and vertical.

In conventional diffuse X-ray scattering with incoherent light, the angular width of a peak is $\sim \lambda/\xi$ where ξ is the average domain size. Such a diffuse peak results from the incoherent sum of the scattering from different domains. With coherent light, the coherent sum of the scattering from a random array of domains results in a speckle pattern modulating the diffuse peak⁶. The angular extent of each 'speckle' is comparable to the width of the central peak in the Fraunhofer diffraction pattern of the collimating pinhole, as this corresponds to a change of λ in the largest path-length difference. This width is $\sim \lambda/L$, where L is the beam diameter. If, instead of diffuse scattering, we consider Bragg diffraction from a highly ordered sample ($\xi \gg L$), the complete peak is the size of a single speckle. We confirmed this by measuring the (111) peak of a perfect silicon crystal which simply reproduced the Fraunhofer diffraction pattern of the collimating pinhole. Because of the limited intensity of our coherent beam, the optimum condition to observe a speckle pattern is obtained with a diffuse peak not too much broader than λ/L .

We chose to measure the diffuse (001) peak from an ordered single crystal of Cu_3Au . The structure of such a crystal consists of a random arrangement of antiphase domains, which arise because of the four equivalent ways in which the Cu_3Au unit can occupy the underlying crystal lattice⁷. Neighbouring domains differ only by a phase shift in the periodic occupation of lattice sites by this unit. By an appropriate heat treatment, the crystal was prepared with an average domain size of a few

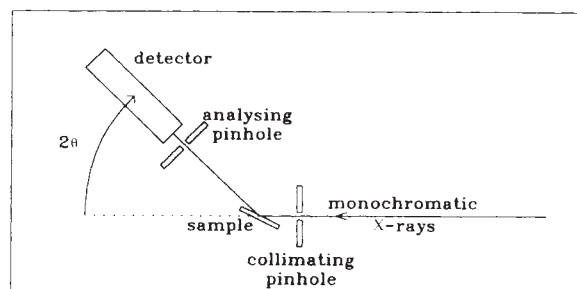


FIG. 1 Schematic representation of the scattering geometry. The angle 2θ is shown; α (not shown) is the corresponding angle in the direction normal to the scattering plane (normal to the page).

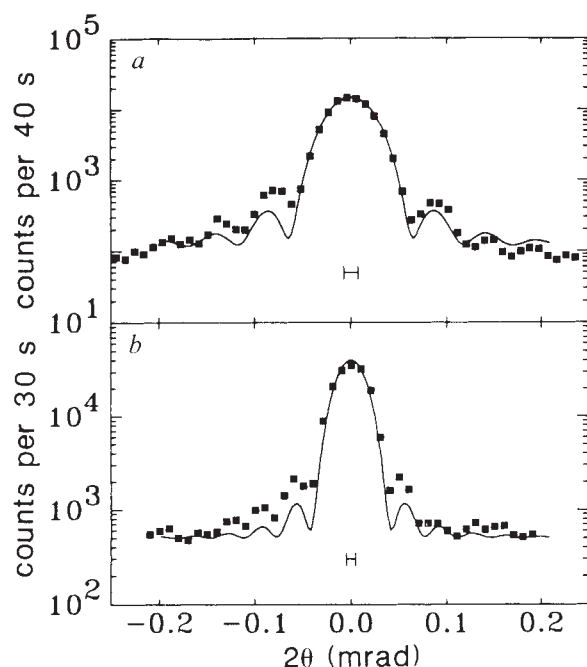


FIG. 2 Fraunhofer diffraction from *a*, 2.5- μm and *b*, 5- μm nominal diameter pinholes. Error bars due to counting statistics are smaller than the symbol size. The solid line represents the calculated diffraction pattern of a circular aperture convolved with the analyser resolution shown by the horizontal bar. A constant background is included in the fit.

hundred nanometers. During our experiment the domain structure was static, as atomic mobility was negligible at room temperature. The (001) peak at $2\theta_B = 23.9^\circ$ had an angular width of 0.3 mrad in 2θ and 0.5 mrad in α . Measurement of the (002) Bragg peak gave widths much sharper than this, confirming that the (001) peak width results from the antiphase domain size rather than from imperfections in the underlying lattice. The scattering path-length differences were smaller than the longitudinal coherence length because of X-ray absorption. In the symmetric reflection geometry, the largest path-length difference

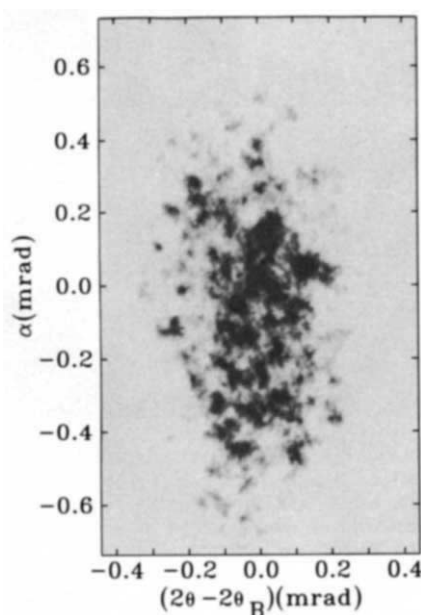


FIG. 3 Photograph of the speckle pattern in the diffuse (001) peak of Cu_3Au . The collimating pinhole is 2.5 μm .

is $2L_z \sin \theta$, where L_z is the penetration depth normal to the surface. For Cu_3Au , $L_z \approx 0.6 \mu\text{m}$.

Figure 3 shows a photograph of the speckle pattern observed from this Cu_3Au crystal. It was obtained by placing film (Kodak DEF) at the analyser position. A 2.5- μm collimating pinhole was used. The diffuse peak is seen to be made up of small intense spots. The angular extent of each spot is comparable to that of the central peak of the Fraunhofer diffraction pattern of the pinhole ($\lambda/L \approx 50 \mu\text{rad}$). A more quantitative measurement is presented in Fig. 4, which shows the scattering patterns obtained using 2.5-, 5- and 50- μm collimating pinholes with 50-, 25- and 100- μm analysing pinholes, respectively. For the first two cases, the observed angular sizes of the intense spots are ~ 50 and $25 \mu\text{rad}$, respectively, again in good agreement with λ/L . The measurement using the 50- μm collimating pinhole is included for comparison. It is equivalent to a conventional incoherent X-ray measurement, because 50 μm is larger than the coherence length of the source.

The observed structure in the (001) peak cannot be explained by a topographic imaging effect, where each of the intense spots arises by independent diffraction from a small region of the illuminated area. This is ruled out because diffraction from a region much smaller than L would spread over an angular region much larger than λ/L . As the observed size of the intense spots is comparable to λ/L , they must arise by diffraction from the full illuminated volume.

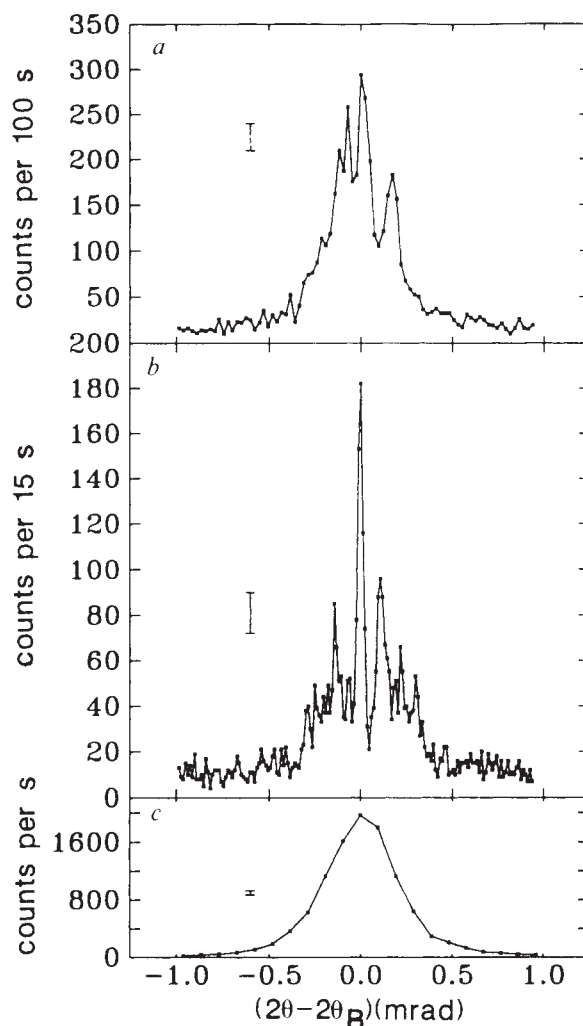


FIG. 4 Speckle patterns measured using *a*, 2.5- μm , *b*, 5 μm and *c*, 50- μm collimating pinholes. The analysing pinholes used were 50, 25 and 100 μm , respectively. Representative error bars are indicated, and the solid lines simply connect the data points. The (001) Bragg angle, $2\theta_B$, is $\sim 23.9^\circ$.

The observation of speckle using X-rays opens up many exciting possibilities for future work. In particular, by measuring the evolution of the speckle pattern in time, we can carry out X-ray intensity fluctuation spectroscopy. This technique not only measures the dynamics of nonequilibrium systems but is also one of the few techniques that can directly measure the dynamics of equilibrium systems. One important class of applications is the study of mass-transport mechanisms on length scales of 1–100 nm in materials such as metal alloys and complex fluids⁸. Of immediate interest are critical phenomena in equilibrium dynamics⁹ and fluctuations during nonequilibrium domain coarsening¹⁰. Another class is the study of phason dynamics in materials with an incommensurate modulation, such as charge-density-wave systems and quasicrystals.

The present measurements using X25 at NSLS give a maximum count rate of 10 counts per s, allowing the study of correlation times down to 0.1 s. Such timescales are relevant to

mass transport over all distances down to the atomic scale in typical solids at temperatures below about half their melting point. New undulator-based facilities at the European synchrotron radiation facility, the advanced photon source and elsewhere will provide an increase of three orders of magnitude in coherent intensity. This will enable studies probing correlation times in the microsecond range and studies of materials that scatter relatively weakly.

Intensity fluctuation spectroscopy does not make full use of the structural information in the speckle pattern. In principle, the speckle pattern contains complete information on the disordered structure producing it⁶. This structure could be imaged by Fourier inversion. In some cases, such as single grain boundaries or dislocations, the speckle pattern will be relatively simple, and one could use it to determine the atomic structure of these defects. □

Received 25 February; accepted 25 June 1991.

1. Pusey, P. N. in *Photon Correlation Spectroscopy and Velocimetry* (eds Cummins, H. Z. & Pike, E. R.) 45–141 (Plenum, New York, 1977).
2. Jakeman, E., Pusey, P. N. & Vaughan, J. M. *Opt. Commun.* **17**, 305–308 (1976).
3. Born, M. & Wolf, E. *Principles of Optics* 396; 511 (Pergamon, Oxford, 1975).
4. Rarback, H., Jacobsen, C., Kirz, J. & McNulty, I. *Nucl. Instrum. Meth.* **A266**, 96–105 (1988).
5. Berman, L. E. & Hart, M. *Nucl. Instrum. Meth.* **A300**, 415–421 (1991).
6. Ludwig, K. F. Jr. *Phys. Rev. Lett.* **61**, 1526 (1988).

7. Warren, B. E. *X-ray Diffraction*, 206–247 (Dover, Mineola, New York, 1990).
8. Lipowsky, R. *Nature* **349**, 475–481 (1990).
9. Hohenberg, P. C. & Halperin, B. I. *Rev. mod. Phys.* **49**, 435–479 (1977).
10. Roland, C. & Grant, M. *Phys. Rev. Lett.* **63**, 551–554 (1989).

ACKNOWLEDGEMENTS. We thank M. Grant for discussions and R. Sweet, Jia Wang, B. Ocko and R. Shannon for their help.

Transition from a uniform state to hexagonal and striped Turing patterns

Q. Ouyang & Harry L. Swinney

Center for Nonlinear Dynamics and Department of Physics,
The University of Texas, Austin, Texas 78712, USA

CHEMICAL travelling waves have been studied experimentally for more than two decades^{1–5}, but the stationary patterns predicted by Turing⁶ in 1952 were observed only recently^{7–9}, as patterns localized along a band in a gel reactor containing a concentration gradient in reagents. The observations are consistent with a mathematical model for their geometry of reactor¹⁰ (see also ref. 11). Here we report the observation of extended (quasi-two-dimensional) Turing patterns and of a Turing bifurcation—a transition, as a control parameter is varied, from a spatially uniform state to a patterned state. These patterns form spontaneously in a thin disc-shaped gel in contact with a reservoir of reagents of the chlorite–iodide–malonic acid reaction¹². Figure 1 shows examples of the hexagonal, striped and mixed patterns that can occur. Turing patterns have similarities to hydrodynamic patterns (see, for example, ref. 13), but are of particular interest because they possess an intrinsic wavelength and have a possible relationship to biological patterns^{14–17}.

The reaction medium is a 2.0-mm-thick polyacrylamide gel disk (25.4-mm diameter), which is sandwiched between two 0.4-mm-thick porous glass disks (Vycor glass, Corning); similar reactors have been described previously^{5,18}. The gel was prepared by the procedure in ref. 7. The gel and glass disks are transparent; the pattern can therefore be detected optically. The outer flat surface of each glass disk is in contact with a chemical reservoir, where reactant concentrations are kept constant and uniform by mixing and a continuous flow of fresh reagents. Components of the chlorite–iodide–malonic acid reaction¹² are distributed in the two reservoirs in such a way that neither is separately reactive. The chemicals diffuse through the porous glass disks into the gel where the reaction occurs. The gel, loaded with a soluble starch indicator (Thiodène, Prolabo), changes colour from yellow to blue with changes in concentration of I_3^-

during the redox reaction. No starch is present in the porous glass; thus concentration changes in the glass disks are not visible. The pattern is monitored in transmitted light (580 nm) with a video camera.

Beyond critical values of the control parameters (chemical concentrations and temperature), patterns emerge spontaneously from an initially uniform background. Initially, after the parameters are switched into a regime where patterns arise,

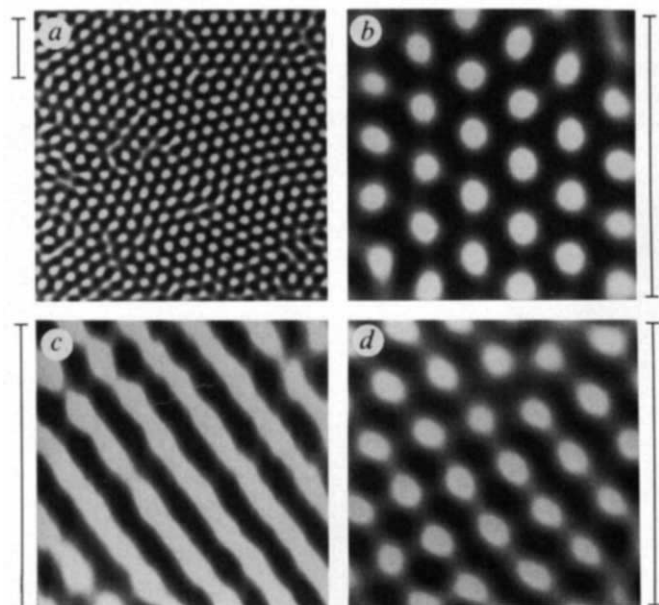


FIG. 1 Stationary chemical patterns formed in a continuously fed laboratory reactor. a, b, Hexagons; c, stripes; d, mixed state. The bar beside each picture represents 1 mm; the reactor is 25 mm in diameter. The concentrations in reservoirs A and B on the two sides of the reactor were: $[I^-]_A^0, [I^-]_B^0, [CH_2(COOH)_2]_B^0$ (in $10^{-3}M$) in a and d, 3.0, 3.0, 13; in b, 3.5, 3.5, 8.3; in c, 5.0, 5.0, 8.3 (for the conditions given, the mixed state in d coexists with the hexagons in a). The parameters common to all observed patterns were: $[NaOH]_A^0 = [NaOH]_B^0 = 3.0 \times 10^{-3}M$, $[Na_2SO_4]_A^0 = [Na_2SO_4]_B^0 = 3.0 \times 10^{-3}M$, $[ClO_2^-]_A^0 = 1.8 \times 10^{-2}M$, $[ClO_2^-]_B^0 = 0$, $[CH_2(COOH)_2]_A^0 = 0$, $[H_2SO_4]_A^0 = 2.0 \times 10^{-3}M$, $[H_2SO_4]_B^0 = 1.0 \times 10^{-2}M$, temperature $5.6^\circ C$.

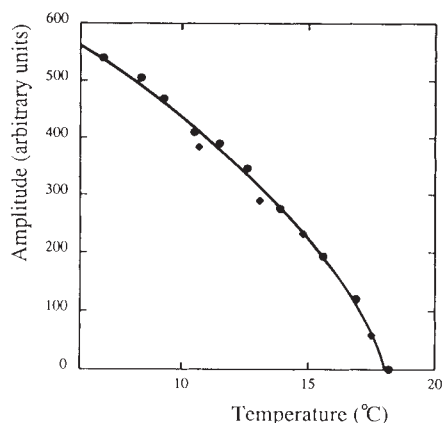


FIG. 2 The transition from a uniform state to a hexagonal Turing pattern. The amplitude was obtained from the modulus of a two-dimensional Fourier transform of a band near the spatial frequency 5.6 mm^{-1} . The points $\bullet$ ($\blacklozenge$) were measured for temperature decreasing (increasing) in steps. Concentrations were the same as in Fig. 1b except that $[\text{ClO}_2]_0 = 2.0 \times 10^{-2} \text{ M}$.

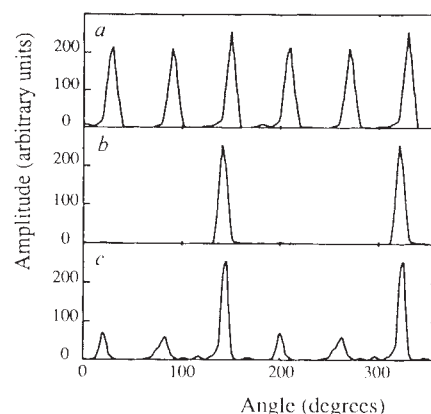


FIG. 3 Angular intensity distributions in spatial frequency space (in a band centred at 5.6 mm^{-1}) for a, hexagons, b, stripes and c, mixed state. Conditions for a, b and c respectively correspond to the conditions in Fig. 1b, c and d.

there are many (~ 100) transient yellow circles growing in a blue background. Within an hour these patterns slowly stop propagating and break up into yellow dot patterns which evolve more slowly. The system approaches a nearly stationary state with domains of hexagonal patterns (Fig. 1b) separated by grain boundaries (Fig. 1a) which move very slowly, typically only 0.2 mm per day.

The transition from the uniform state to a hexagonal pattern was studied with temperature as a control parameter and other parameters held fixed: the hexagonal pattern emerged as the temperature was lowered, as Fig. 2 illustrates. Within the experimental resolution there was no hysteresis; the transition occurred at $18.0 \pm 0.5^\circ \text{C}$ for both increasing and decreasing temperature. The system showed a critical slowing down at temperatures near the transition: the relaxation time became days, whereas far from the transition (for example at 5.6°C) the system relaxed within 3 hours. These observations suggest that the transition is continuous, but a small discontinuity, as expected from theory (to be discussed), would not be observable with our resolution.

Stripes rather than hexagons were observed at high iodide or low malonic acid concentration. The stripes (Fig. 1c) were stationary, in contrast to the widely studied travelling wave patterns¹⁻⁵. Striped patterns were maintained for days without much change except for a very slow motion of the grain boundaries separating different domains.

The wavelengths of the hexagonal and striped patterns, determined from spatial fast Fourier transforms, varied continuously from 0.14 to 0.33 mm with changes in the control parameters. The wavelength is an intrinsic property of the reaction-diffusion system, not a consequence of finite size of the system. This intrinsic wavelength distinguishes Turing patterns from other well known nonequilibrium structures such as convection rolls or Taylor vortices. The wavelength was especially sensitive to the sulphuric acid concentration in reservoir A; for example a decrease of sulphuric acid concentration in reservoir A by 33% led to a 45% increase in wavelength of the hexagons and stripes.

A distinct stable mixed state was observed as well as the pure hexagons and stripes. Mixed states are discernible in photographs, as in Fig. 1d, and can be quantified in graphs of the angular distribution of amplitude (Fig. 3). Pure hexagons have six peaks of equal amplitude separated by 60° (Fig. 3a); pure stripes have two peaks separated by 180° (Fig. 3b); and the observed mixed state has six peaks separated by 60° , with one large peak followed by two small, one large, and two small (Fig. 3c).

Two-dimensional chemical patterns similar to those we have observed have been found in numerical simulations of reaction-

diffusion models with the kinetics given by the Brusselator¹⁹ or variants²⁰ or by an activator-inhibitor²¹ model. Pattern formation in two-dimensional systems has also been extensively studied using general equations that describe the amplitude of the patterns²². In the absence of specific symmetries, which is the usual situation in chemistry, the initial transition from a uniform state to hexagons is discontinuous²³⁻²⁵. But in practice the sizes of the discontinuity and hysteresis range are usually small, and the transition shares most features with a continuous transition. Simulations by V. Dufiet and J. Boissonade (personal communication) on a two-species reaction-diffusion model²⁶ reveal a transition with a hysteresis that is only 0.1% of the control parameter, too small to observe with our experimental resolution.

Our experiments show evidence of a transition to Turing patterns that can be maintained indefinitely in a well-defined nonequilibrium state. We have neglected possible structure in the direction of the chemical gradient; such structure might exist, as the thickness of the gel is 6–10 times as great as the wavelength of the pattern. Walgraef *et al.*²⁷ have shown that in three-dimensional systems with no imposed gradients, the most likely selected patterns should be body-centred cubic, hexagonal prisms and stripes, all of which have been found in simulations (A. DeWit *et al.*, personal communication). Only the body-centred cubic pattern has structure in three dimensions; the patterns we have observed, hexagons and stripes, have no three-dimensional structure.

Further experiments are under way to search for the other types of patterns and to study the transitions between patterns, multi-stability, the dynamics of defects and the role of three-dimensionality. The development of continuously fed spatially extended reactors such as the one described here opens a wide range of problems to systematic laboratory study. $\square$

Received 22 April; accepted 14 June 1991.

1. Zaikin, A. N. & Zhabotinskii, A. M. *Nature* **225**, 535–537 (1970).
2. Field, R. J. & Berger, M. (eds) *Oscillations and Traveling Waves in Chemical Systems* (Wiley, New York, 1985).
3. Winfree, A. T. *When Time Breaks Down* (Princeton University Press, 1987).
4. Tyson, J. J. & Keener, J. P. *Physica D* **32**, 327–361 (1988).
5. Noszticzius, Z., Horsthemke, W., McCormick, W. D., Swinney, H. L. & Tam, W. Y. *Nature* **329**, 619–620 (1987).
6. Turing, A. M. *Phil. Trans. R. Soc. Lond.* **B327**, 37–72 (1952).
7. Castets, V., Dulos, E., Boissonade, J. & de Kepper, P. *Phys. Rev. Lett.* **64**, 2953–2956 (1990).
8. Boissonade, J., Castets, V., Dulos, E. & de Kepper, P. *ISNM* **97**, 67–77 (Birkhäuser, Basel, 1991).
9. De Kepper, P., Castets, V., Dulos, E. & Boissonade, J. *Physica D* **49**, 161–169 (1991).
10. Boissonade, J. *J. Phys.* **49**, 541–546 (1988).
11. Lengyel, I. & Epstein, I. R. *Science* **251**, 650–652 (1991).
12. De Kepper, P., Boissonade, J. & Epstein, I. R. *J. phys. Chem.* **94**, 6525–6536 (1990).
13. White, D. B. *J. Fluid Mech.* **191**, 247–286 (1988).

14. Nocolis, G. & Prigogine, I. *Self-Organization in Nonequilibrium Systems* (Wiley, New York, 1977).
15. Meinhardt, H. *Models of Biological Pattern Formation* (Academic New York, 1982).
16. Harrison, L. G. *J. theor. Biol.* **125**, 369–384 (1987).
17. Murray, J. D. *Mathematical Biology* (Springer, Berlin, 1989).
18. Kshirsagar, G., Noszticzus, Z., McCormick, W. & Swinney, H. L. *Physica* **D49**, 5–12 (1991).
19. Borkmans, P., Dewel, G. & De Wit, A. in *Seeds: Genesis of Natural and Artificial Forms*, 62–71 (Le Biopole Végétal, Amiens, 1991).
20. Lacalli, T. C., Wilkinson, D. A. & Harrison, L. C. *Development* **104**, 105–113 (1988).
21. Haken, H. & Olbrich, H. *J. math. Biol.* **6**, 317–331 (1978).
22. Mannerville, P. *Dissipative Structures and Weak Turbulence* (Academic, Boston, 1990).
23. Busse, F. H. *J. Fluid Mech.* **30**, 625–649 (1967).
24. Pismen, L. M. *J. chem. Phys.* **72**, 1900–1907 (1980).
25. Malomed, B. A. & Tribel'skii, M. I. *Sov. Phys. JETP* **65**, 3013–3016 (1987).
26. Schnackenberg, J. J. *J. theor. Biol.* **81**, 389–400 (1979).
27. Walgraef, D., Dewel, G. & Borkmans, P. *Phys. Rev. A* **A21**, 397–404 (1980); *Adv. chem. Phys.* **49**, 311–355 (1982).

ACKNOWLEDGEMENTS. We thank J. Boissonade and V. Duflet for numerical simulations and for comments on the manuscript, and A. Armeodo, P. Borchmans, P. de Kepper, G. Dewel, A. de Wit, W. McCormick, D. Vigil and D. Walgraef for discussions. This work was supported by the US Department of Energy Office of Basic Energy Sciences and BP Venture Research.

High turnover rates of dissolved organic carbon during a spring phytoplankton bloom

David L. Kirchman*, Yoshimi Suzuki†, Christopher Garside‡ & Hugh W. Ducklow§

* College of Marine Studies, University of Delaware, Lewes, Delaware 19958, USA

† Geochemical Laboratory, Meteorological Research Institute, Nagamine 1-1 Tsukuba-shi, Ibaraki, 305 Japan

‡ Bigelow Laboratory, West Boothbay Harbor, Maine 04575, USA

§ Horn Point Environmental Laboratory, University of Maryland, Cambridge, Maryland 21613, USA

OCEANIC dissolved organic carbon (DOC) is one of the Earth's largest carbon reservoirs, but until recently its role in the carbon cycle has been neglected. New methodology¹, however, has led to larger estimates of DOC concentrations and also to renewed interest in the biochemical lability of DOC². Previous work found that the mean age of DOC in the surface ocean was >1,000 years³. To examine the lability of DOC in greater detail, we have conducted experiments to estimate DOC turnover rates in the upper ocean. We directly observed rapid DOC turnover by bacterioplankton during the spring phytoplankton bloom in the North Atlantic ocean. Potential turnover rates, measured in 0.8- μ m filtered samples, ranged from 0.025 to 0.363 per day, and were consistent with bacterial biomass production and uptake of dissolved nitrogen (NH_4^+ , NO_3^- and urea). Our results indirectly suggest that cycling of dissolved organic nitrogen (DON) differs from that of DOC. The high estimates of DOC concentrations and turnover rates repeated here, if found to be general, would seem to demand changes in models⁴ of carbon cycling and of the ocean's role in buffering increases in atmospheric CO_2 .

Our experiments were conducted as part of the Joint Global Ocean Flux Study (JGOFS), which had the objective of examining the 1989 spring phytoplankton bloom in the North Atlantic Ocean (47° N; 18° W) and its role in the ocean carbon cycle⁵. The bloom started in late April, as indicated by increases in chlorophyll *a* and decreases in NO_3^- and total inorganic carbon. It continued until at least early June, sustained by ammonium recycling and residual NO_3^- . To examine the possible rate of degradation of DOC, we gravity-filtered surface sea water through Nuclepore filters with pore sizes of 0.8 μ m (diameter 142 mm), added an equal volume of the same water filtered through 0.22- μ m Millipore filters (which remove all microorganisms) and then incubated this mixture in the dark at ambient temperature (~14 °C). All filters were rinsed with distilled water and sample water before use. The 0.8- μ m filtrate contained nearly the entire heterotrophic bacterial assemblage and only 10% of chlorophyll *a*; that is, initially it did not contain sources of DOC (phytoplankton and zooplankton) although it retained

those organisms (bacteria) thought to be mainly responsible for DOC oxidation.

Bacterial abundance (measured by direct counts⁶) increased with time (Fig. 1a), because 0.8- μ m filtration excludes many grazers of bacteria, and dilution further minimizes bacterivore activity^{7,8}. After 4.2 days, the bacterial abundance decreased because bacterivorous microflagellates increased from initially very low numbers to populations that had a considerable effect on the bacteria (Fig. 1a). Accompanying the initial increase in bacterial abundance, there was a decrease in DOC concentrations (measured by high-temperature catalytic oxidation¹) in these <0.8- μ m filtrates (Fig. 1b). For the first day, the apparent rate of DOC degradation was high, ~0.2 per day (Table 1; rates are first-order decay constants). The rate decreased during the incubation period (Fig. 1b), and rates were lower when calculated over intervals greater than 2 days (Table 1). The lowest turnover rate of DOC was 0.025 ± 0.005 per day, estimated after 4.2 days in an incubation lasting 11 days. At least some of the apparent decrease in turnover rates was probably related to the increase in bacterivore activity. Grazing by bacterivores lowers bacterial abundance and releases DOC⁹, thus slowing its net depletion. Even so, our lowest turnover rate is still greater than those estimated in previous studies^{10,11} similar to ours.

The high DOC turnover rates did not result from the release of labile DOC during filtration. Concentrations of DOC before and after filtration were similar, indicating that any DOC contamination was small. Also, the fraction of DOC consumed in these experiments (23–42%; Table 1) is too high to be explained by degradation of the small amount (nM) of DOC perhaps released by cell lysis during filtration. Although our estimates at the end of these experiments did approach the low DOC concentrations measured by the wet-combustion method (*in situ* concentrations measured by this method¹² are about half those measured by high-temperature catalytic oxidation), our incubations probably did not last long enough to determine what fraction of the DOC was highly refractory with low turnover rates. Nonetheless, the lowest initial DOC turnover rate did coincide with the lowest initial concentration (31 May; Table 1). In short, our results indicate that DOC is more labile in surface waters than indicated by previous estimates of DOC age and turnover^{3,10,11}.

Bacterial activity (and possibly that of other microorganisms) was essential to deplete the DOC concentrations in the 0.8- μ m filtrates. The concentrations of DOC did not decrease when sea water was filtered through 0.22- μ m Millipore filters, which remove all microorganisms (Fig. 1b), indicating that the decrease was not due to adsorption on the container walls or abiotic oxidation.

The nitrogen required for the observed bacterial growth was apparently supplied by NH_4^+ , NO_3^- and urea. The concentrations of these nitrogenous compounds decreased during periods when

TABLE 1 Summary of dissolved organic carbon (DOC) turnover rates calculated from DOC depletion in <0.8- μ m filtrates

Start date (1989)	Incubation period (d)	Initial DOC (μ M)	Fraction consumed (%)	Turnover rate (d^{-1})*	n
25 May	0–1.0	196 ± 8		0.230 ± 0.020	4
	1.0–4.2	155 ± 3		0.048 ± 0.003	6
	4.2–11	134 ± 1	42	0.025 ± 0.005	4
28 May	0–1.0	178 ± 2		0.363 ± 0.290	3
	1.3–4.0	146 ± 2	28	0.044 ± 0.009	4
31 May	0–3.0	136 ± 3	23	0.087 ± 0.017	5

* Turnover rates were calculated by applying first-order kinetics to incubation periods during which graphs of $\ln$ (DOC concentration) against time were linear. The rate is then the slope of that line ($\pm$ standard error) calculated with *n* time points.

bacterial abundance increased and DOC concentrations decreased in $<0.8\text{-}\mu\text{m}$ filtrates (Fig. 1c). We therefore have a direct observation of the use of NO_3^- and urea by the 'bacterial size fraction' (NH_4^+ use by bacterial assemblages is well known already¹³). The increase in bacterial nitrogen was less than or equal to the decrease in NH_4^+ , NO_3^- and urea concentrations, implying that little growth was supported by DON during these experiments. Concentrations of NH_4^+ increased late in the experiments (after 4.2 days; Fig. 1c), probably because of grazing activity, not DON mineralization. This NH_4^+ increase covaried with a decrease in bacterial abundance and an increase in microflagellate abundance (Fig. 1).

To determine whether rates of DOC degradation, N uptake and bacterial biomass production were consistent, we calculated carbon growth efficiencies by comparing both the increase in bacterial biomass (for cell-based growth efficiencies) and the decrease in nitrogen concentrations (for nitrogen-based efficiencies) with the decrease in DOC concentrations (Table 2). The cell-based and N-based carbon growth efficiencies were $1.6\text{--}9.0\%$ and $26 \pm 11\%$, respectively (Table 2). Using a similar approach, others¹⁴ also have found cell-based carbon growth efficiencies to be as low as 1%, but most estimates are $>20\%$ (see, for

example, ref. 15). The difficulties in measuring bacterial carbon do not explain the low cell-based efficiencies. In contrast, N-based growth efficiencies were calculated from changes in concentrations of NH_4^+ , NO_3^- and urea, which are easily measured. These growth efficiencies are similar to most previous estimates¹⁵ and indicate that the high rates of DOC depletion are consistent with measured disappearance of inorganic nitrogen and urea.

To examine how the potential turnover rates of DOC calculated from these experiments compare with processes in the ocean itself, we made a first-order estimate of *in situ* DOC turnover from the following expression:

$$\text{DOC turnover rate} = \frac{\text{Bacterial production}}{(\text{DOC concentration}) \times (\text{growth yield})}$$

Bacterial production was estimated from rates of ^3H -thymidine and ^3H -leucine incorporation in samples taken directly from depth profiles to 200 m. These measurements include all microorganisms, and microbial processes were not intentionally disrupted, unlike our approach to estimating potential DOC turnover (Table 1). During our study, bacterial production in the mixed layer was $\sim 0.4\text{ }\mu\text{M C}$ per day, and the DOC concentration was $\sim 200\text{ }\mu\text{M C}$. The estimated *in situ* rate of DOC turnover ranges from 0.006 to 0.1 per day, assuming growth efficiencies range from 35 to 1.6% (Table 2).

These estimates of *in situ* turnover rates are substantially lower than the potential rate if the growth efficiency is assumed to be $>30\%$ (Table 2); such growth efficiencies have frequently been observed by others¹⁵. The difference in turnover rates may be due to stimulation of bacterial activity in the $<0.8\text{-}\mu\text{m}$ filtrates used to estimate the potential rates. The bacterial cell volume (measured microscopically) was three times as great and the incorporation rates per cell of thymidine and leucine 15 times as great in the $<0.8\text{-}\mu\text{m}$ filtrates as in the initial sample or in *in situ* measurements. Because of these large changes in the bacterial assemblage, the high potential turnover rates should be viewed with caution.

The high turnover rates do, however, help to explain the large changes in DOC concentrations (± 10 to $50\text{ }\mu\text{M}$) that we observed *in situ* during the spring bloom in the North Atlantic. It is difficult to distinguish temporal from spatial variability in the open sea. If spatial variability is ignored, however, these changes in DOC concentrations require that $>100\%$ of primary production (as measured by ^{14}C) is occasionally routed through the DOC reservoir. Even before development of the new methods of measuring DOC, Williams¹⁶ suggested that 50–60% of primary production goes through a DOC phase. We cannot calculate the turnover rate of DOC in the oceans and its relationship to primary production precisely, because the mechanisms

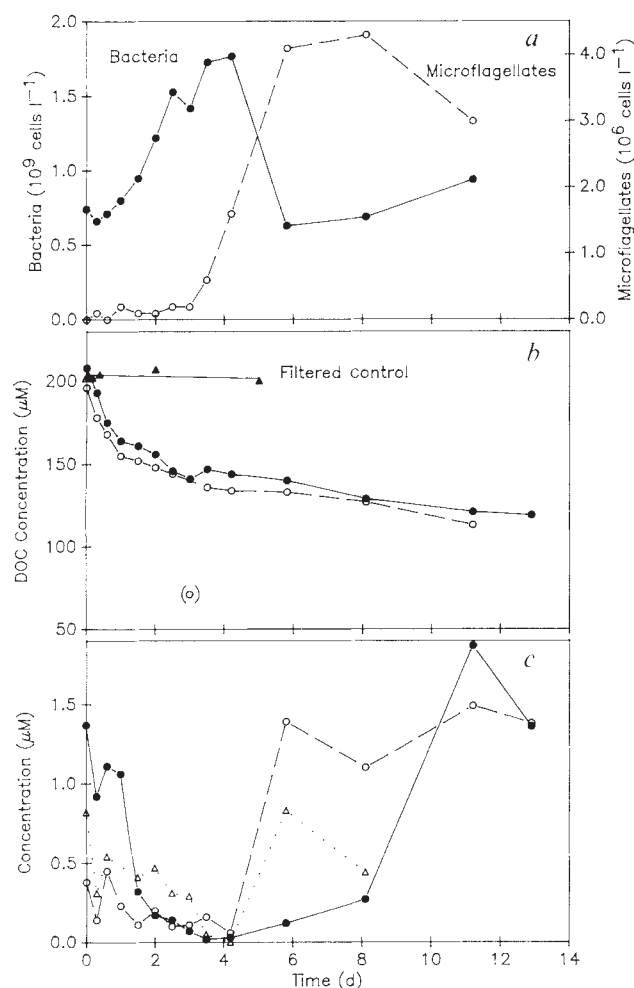


FIG. 1 Abundance of *a*, bacteria (●) and microflagellates (○); *b*, dissolved organic carbon (○), total organic carbon (●) and filtered control (▲); *c*, concentrations of nitrogen compounds (○, ammonium; ●, nitrate; △, urea), after $0.8\text{-}\mu\text{m}$ filtration, 25 May 1989 in the North Atlantic. For the filtered control, taken on 29 May, surface water was first filtered through $0.22\text{-}\mu\text{m}$ Millipore filters to remove all microorganisms before incubation. Bacterial abundance was measured by epifluorescence microscopy after staining with acridine orange⁶. Dissolved organic carbon was measured by the high-temperature oxidation method⁴. The nitrogen compounds were measured by standard wet chemistry methods²⁹.

TABLE 2 Summary of carbon growth efficiencies during DOC depletion experiments with $<0.8\text{-}\mu\text{m}$ filtrates

Start date (1989)	Incubation time (days)	Biomass production ($\mu\text{M C}$)	DOC depletion ($\mu\text{M C}$)	Total N-uptake* ($\mu\text{M N}$)	Carbon growth efficiency†	
					Cell (%)	N (%)
25 May	4.2	1.2–4.9‡	62	3.3	1.9–7.9	24
28 May	3.0	0.9–3.6	40	1.6	2.2–9.0	18
31 May	3.0	0.5–2.1	31	2.4	1.6–6.8	35

* Total N uptake calculated from disappearance of NH_4^+ , NO_3^- and urea over time.

† Carbon growth efficiencies were defined as follows: cell-based carbon growth efficiency = (biomass production)/(DOC depletion); N-based carbon growth efficiency = (total N-uptake $\times 4.5$)/(DOC depletion), where 4.5 is the C:N ratio for bacteria²⁷.

‡ Range reflects uncertainty in converting bacterial biovolumes (measured microscopically) to biomass. Conversion factors used were $140\text{ g C }\mu\text{m}^{-3}$ (ref. 27) and $500\text{ fg C }\mu\text{m}^{-3}$ (ref. 28).

of DOC production are not well understood and because of the possible variation with location. Nonetheless, our high potential rates and large gradients in DOC concentrations indicate that a large fraction of DOC (20–40%) can turn over in days and that the DOC flux is large relative to primary production. Such large, short-term fluxes cannot be explained with current concepts in plankton ecology.

Previous work had indicated that turnover rates of components of the oceanic DOC reservoir vary from less than a day to more than 1,000 years. Compounds such as dissolved free amino acids can have turnover times of minutes¹⁷, but these compounds are <10% of total DOC¹⁸. Williams and Druffel³ calculated that 56% of surface DOC was <30 years old and only the remaining 44% was very old (>6,000 years). Duursma's work¹⁹ suggests that DOC changes seasonally in the North Atlantic, implying that a large fraction turns over in less than a year. Our results suggest even higher rates, with roughly 20–40% turning over in days (Table 1 and changes measured *in situ*).

Our high estimates of potential DOC turnover (Table 1) may be related to the spring phytoplankton bloom from which we sampled. In the North Sea, Duursma¹⁹ found that DOC concentrations (measured by wet combustion) were highest in spring during a phytoplankton bloom and decreased to a third by winter. Data from April and September 1958 in the North Atlantic Ocean suggest a similar seasonal variation in open-ocean concentrations. Our high potential rates may therefore be due to high DOC concentrations during spring blooms. In any event, these high rates suggest that DOC concentrations can vary in days rather than in seasons.

Turnover of DON, in contrast, may differ greatly from that of DOC, at least during phytoplankton blooms. The simultaneous uptake of DOC and of inorganic nitrogen and urea implies that DON (other than urea and free amino acids) was not assimilated extensively by marine microorganisms in these experiments. The amino acid pool is assimilated rapidly in <0.8- μ m filtrates²⁰, as is urea, but their concentrations make up <10% of DON¹⁸. In one experiment (31 May), the ammonium concentration did increase initially, suggesting that DON was being used, but other studies^{20,21} have found that mineralization of dissolved free amino acids alone can explain much of this ammonium regeneration. That DOC and DON have different cycles is also suggested by the seasonal change in the DOC:DON ratio in the North Sea¹⁹ and the lack of covariance between DOC and DON in the Pacific Ocean²², although these previous studies used the wet-combustion method.

Differences in DOC and DON cycling may be greatest during phytoplankton blooms because the dissolved material then produced is mostly DOC with little or no nitrogen, such as mono- and polysaccharides²³. Uptake of this DOC would require simultaneous uptake of inorganic nitrogen²⁴. Parsons *et al.*²⁵ suggested that heterotrophic bacteria assimilated NO₃⁻ in experiments with large glucose additions. High turnover of DOC explains the unusual heterotrophic uptake of NO₃⁻ that we observed in these experiments and also in ¹⁵N measurements with *in situ* microbial assemblages. The disappearance of inorganic nitrogen balanced the increase in bacterial nitrogen in our experiments, suggesting that the turnover of DON, unlike DOC, was low. To examine possible differences in DOC and DON cycling in greater detail, the methodology²⁶ for measuring DON concentrations in sea water needs to be tested. □

8. Ducklow, H. W. & Hill, S. M. *Limnol. Oceanogr.* **30**, 239–259 (1985).
9. Taylor, G. T. *et al. Mar. Ecol. Prog. Ser.* **23**, 129–141 (1985).
10. Barber, R. T. *Nature* **220**, 274–275 (1968).
11. Ogura, N. *Mar. Biol.* **13**, 89–93 (1972).
12. Williams, P. J. leB. *Chemical Oceanography* (eds Riley, J. P. & Skirrow, G.) 301–363 (Academic, London, 1975).
13. Wheeler, P. A. & Kirchman, D. L. *Limnol. Oceanogr.* **31**, 998–1009 (1986).
14. Meyer, J. L. *et al. Microbiol. Ecol.* **13**, 13–29 (1987).
15. Bjørnsen, P. K. *Mar. Ecol. Prog. Ser.* **30**, 191–196 (1986).
16. Williams, P. J. leB. *Kiel. Meeresforsch.* **5**, 1–28 (1981).
17. Fuhrman, J. A. *Mar. Ecol. Prog. Ser.* **66**, 197–203 (1990).
18. Sharp, J. H. *Nitrogen in the Marine Environment* (eds Carpenter, E. J. & Capone, D. G.) 1–35 (Academic, New York, 1983).
19. Duursma, E. K. *Neth. J. Sea Res.* **1**, 3–190 (1961).
20. Keil, R. G. & Kirchman, D. L. *Mar. Ecol. Prog. Ser.* **73**, 1–10 (1991).
21. Kirchman, D. L. *et al. Deep-Sea Res.* **36**, 1763–1776 (1989).
22. Jackson, G. A. & Williams, P. M. *Deep-Sea Res.* **32**, 223–235 (1985).
23. Ittekkot, V. *et al. Mar. Ecol. Prog. Ser.* **4**, 299–305 (1981).
24. Kirchman, D. L. *et al. Limnol. Oceanogr.* **35**, 1258–1266 (1990).
25. Parsons, T. R. *et al. Mar. env. Res.* **4**, 229–242 (1980/1981).
26. Suzuki, Y. *et al. Mar. Chem.* **16**, 83–97 (1985).
27. Nagata, T. & Watanabe, Y. *Appl. env. Microbiol.* **56**, 1303–1309 (1990).
28. Bratbak, G. *Appl. env. Microbiol.* **53**, 1488–1493 (1985).
29. Whittledge, T. E. *et al. Automated Nutrient Analyses in Seawater, Brookhaven Natl. Lab. Publ. No. 51398* (Department of Energy and Environment, Upton, New York, 1981).

ACKNOWLEDGEMENTS. We thank H. Quinby for technical assistance and T. Nagata, P. J. leB. Williams and J. R. Toggweiler for comments on the manuscript. This research was supported by the NSF.

A lunar meteorite found outside the Antarctic

D. H. Hill*, W. V. Boynton*† & R. A. Haag‡

* Lunar & Planetary Laboratory, University of Arizona, Tucson, Arizona 85721, USA

† Department of Planetary Sciences, University of Arizona, Tucson, Arizona 85721, USA

‡ Robert A. Haag Meteorites, PO Box 27527, Tucson, Arizona 85726, USA

OUR knowledge of the Moon's surface composition has come from samples returned by the Apollo and Luna missions, and from eleven lunar meteorites, all of which were discovered in Antarctica^{1,2}. Here we report the discovery of a new lunar meteorite, Calalong Creek, in a desert region of Australia which is analogous to Antarctica in its ability to preserve meteorites of different types³. On the basis of a diagnostic Fe/Mn ratio of 73–78, and other element abundances, we conclude that Calalong Creek is a lunar breccia, containing both highland and mare materials. Whereas the Apollo and Luna missions selectively sampled only 5% of the lunar crust, lunar meteorites should provide a random sample⁴; nevertheless there has been some concern that the Antarctic meteorite population may be biased in some way⁵. Calalong Creek will add to our understanding of lunar petrology, and as the first non-Antarctic lunar meteorite, may also shed new light on the transfer of impact ejecta from the Moon to the Earth.

Calalong Creek is named after the Aborigine word 'Kalkal-lupilinguta', meaning 'seven sisters went up into the sky, chased by the Moon' (F. Wlotzka and A. Bevan, personal communication). It is a single stone, ~3 cm in diameter, weighing 19 g, with 100% fusion crust covering its exterior. The recovered sample has a caramel-coloured glassy fusion crust with an area containing small vesicles. Sawcuts of the sample reveal a micro-breccia with submillimetre white clasts embedded in a dark-grey matrix (Fig. 1) resembling the first Antarctic lunar meteorite, ALHA81005, although the clasts are much smaller. Microscopic observation of bulk pieces reveals a variety of components: angular white clasts, 'sugary' clasts, white crystalline plagioclase, yellow spinel, pyroxenes and dark grains. The remaining matrix is dark and fine-grained with regions of glassy matrix which appear black and 'spongy' with tiny vesicles. In thin section it resembles lunar clast-laden impact-melt breccias of the Stoffer *et al.*⁶ classification.

Representative chips of Calalong Creek were removed for analysis by instrumental neutron activation analysis (INAA). We found that the meteorite was very compact and consolidated,

Received 15 January; accepted 2 July 1991.

1. Sigumura, Y. & Suzuki, Y. *Mar. Chem.* **24**, 105–131 (1988).
2. Toggweiler, J. R. *Nature* **334**, 468 (1988).
3. Williams, P. M. & Druffel, E. R. M. *Nature* **330**, 246–248 (1987).
4. Toggweiler, J. R. *Productivity of the Ocean: Present and Past* (eds Berger, W. H. *et al.*) 65–84 (Wiley, New York, 1989).
5. Ducklow, H. W. *Oceanogr. Mag.* **2**, 4–8 (1989).
6. Hobbie, J. E. *et al. Appl. env. Microbiol.* **33**, 1223–1228 (1977).
7. Landry, M. R. & Hassett, R. P. *Mar. Biol.* **16**, 283–288 (1982).

TABLE 1 Representative element concentrations*

	Mass	Na	Mg (%)	Al (%)	K	Ca (%)	Sc	Ti	V	Cr	Mn	Fe (%)	Co	Ni	La	Eu	Lu	Hf	Th
Calcalong Creek 9,2,1	18.1 mg	3,590	—	11.2	2,040	11.5	23.1	4,400	54	1,339	1,133	8.49	23.6	370	22.3	1.21	1.10	7.8	4.7
Calcalong Creek 9,2,2	6.2 mg	3,670	3.1	10.7	1,920	11.5	22.3	5,900	65	1,195	1,068	8.40	25.6	330	21.1	1.33	1.09	7.6	4.4
Millbillillie	3.6 mg	3,550	3.1	6.66	330	10	32.0	5,000	83	3,070	4,840	17.2	10	—	3.2	0.66	0.39	3.9	—

* Concentrations are given in p.p.m. except where noted.

Elemental ratios in samples of Calcalong Creek, lunar samples and samples from other parent bodies. Experimental uncertainties are: 1–3% for Na, Al, K, Sc, Cr, Mn, Fe, La, Sm, Eu, Dy, Ho and Yb; 5–15% for Mg, Ca, Ti, V, Co, Ni, Ce, Tb, Lu, Hf, Th and U. (The smaller Calcalong Creek sample generally had better determinations because of more favourable counting conditions.)

making separation of individual clasts difficult. We analysed an 18-mg bulk sample (9, 2, 1) and a 6.2-mg bulk sample (9, 2, 2) of Calcalong Creek, and a 3.6-mg chip of Millbillillie for comparison, as the new meteorite was collected along with other Millbillillie stones, and it was possible that it was a new lithology of Millbillillie rather than a previously unknown meteorite. Recognizing that the meteorite could be precious, we analysed small samples to preserve as much material as possible for future consortium studies. The samples were irradiated twice at the University of Arizona TRIGA reactor with a neutron fluence of $3 \times 10^{13} \text{ cm}^{-2}$ and $2.5 \times 10^{15} \text{ cm}^{-2}$. Although the fluence was lower than that normally used for small samples, it nevertheless generated data with sufficient precision to show that the samples are clearly lunar in origin. The results are presented in Table 1 and Fig. 2a.

In general, it is easy to demonstrate that a meteorite is not from a particular parent body. It is sufficient to show that certain isotope or element ratios are inconsistent with what is known about the parent body. In this case, however, we not only need to show that the meteorite is consistent with a lunar origin, but also to show that no other parent body is a plausible source. It will be shown that the elemental abundance patterns of Calcalong Creek are similar to those seen in other lunar samples and are controlled by such a wide variety of variables that no other parent body is a plausible source.

Laul and Schmitt⁷ reported on the use of the diagnostic Fe/Mn ratio to distinguish the moon and other planetary bodies. Having

the proper Fe/Mn ratio is a necessary, but not sufficient, condition of lunar origin. The ratio is the result of the refractory-volatile fractionation that occurred in the solar nebula before the parent body was formed and metal-silicate fractionation that occurred during its formation. After formation of the parent body, the Fe/Mn ratio in the mantle can be changed as a result of core formation, in which iron but not manganese is removed. All later geochemical processes, however, are relatively ineffective at fractionating iron from manganese, as the igneous partitioning behaviour of these elements is very similar. Consequently iron and manganese are found to be tightly correlated in samples from the same parent body.

All samples returned from the Moon have Fe/Mn ratios of ~75 (ref. 7); samples from most other parent bodies have different ratios. Our values of 73 and 78 fall within the lunar range, whereas Millbillillie and other achondrites are well outside (Table 2). High cobalt and low chromium in the sample also preclude classification of the new meteorite as eucritic, as shown by Warren and Kallemeyn⁸ in the case of lunar meteorite EET87521. Volatile-refractory ratios such as K/La (Table 2) for Calcalong Creek also argue for lunar provenance.

Incompatible elements are important in establishing that this meteorite is lunar. These elements are not easily accommodated by common rock-forming minerals; they therefore become concentrated in the melt phase during partial melting or remain in the residual liquid during fractional crystallization. The abundance of each element in the melt is determined by the fraction

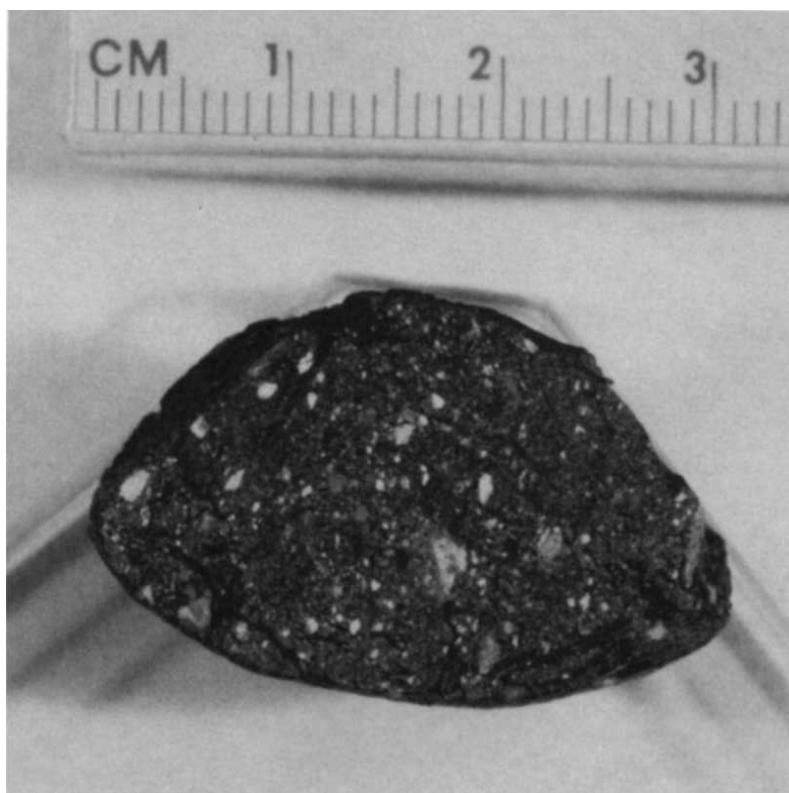


FIG. 1 Hand specimen of the Calcalong Creek meteorite. The 19-g sample was recovered with complete fusion crust. There was little weathering, indicating it must be a relatively recent fall. It has been sawed to reveal submillimetre, white angular clasts in a dark grey matrix (long axis, 3 cm).

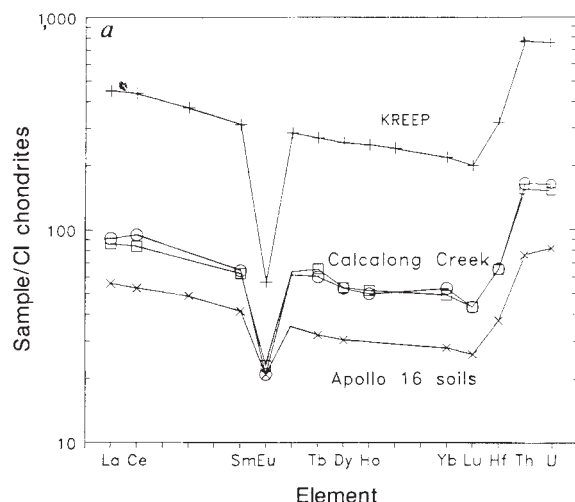
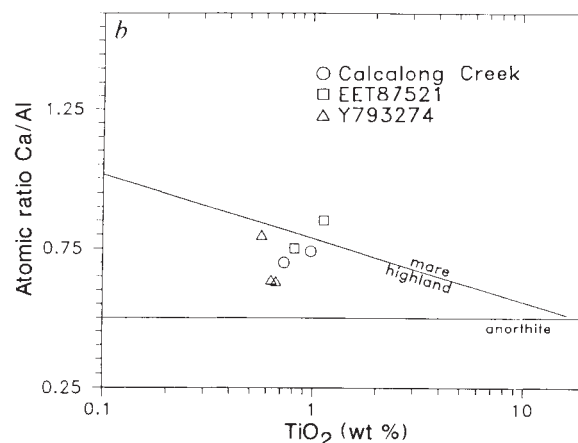


FIG. 2 *a*, Incompatible-element abundances relative to CI chondrites. (KREEP data are from ref. 15 and 16; Apollo-16 soils are from ref. 7. *b*, Calcalong



Creek and other lunar meteorites plotted on Wood's highland/mare diagram of Ca/Al against Ti (ref. 9), illustrating the sampling effect of different abundances of highland and mare components in breccias. The line of demarcation is defined by polymict lunar breccias 15558 and 79135. Data for Y793274 are from refs 5 and 11; EET87521 are from refs 8 and 11.

melted and by the bulk partition coefficient of that element. The bulk mineral/liquid partition coefficient is the average of the coefficients in the individual minerals, weighted by the mass fraction of the minerals present in the source region.

Abundances of rare-earth elements (REE) and other incompatible elements in Calcalong Creek and several other lunar samples are shown in Fig. 2*a*. It is evident that these samples have similar abundance ratios, although the absolute concentrations differ. This abundance pattern is common in many lunar samples and is characteristic of a component called KREEP (for potassium, REE and phosphorus). The lunar KREEP pattern, which is ubiquitous on the Moon, would be impossible to produce on another body unless the same proportions of minerals were present in the source region and the same degree of melting or crystallization were achieved.

It is clear that Calcalong Creek is lunar. The processes and conditions that control the ratios of the elements discussed above

include nebular volatility fractionation, planetary core formation, oxygen fugacity, the composition of the KREEP source material and the degree of crystallization or melting. It is most unlikely that these conditions would occur elsewhere in the Solar System.

The abundance of the incompatibles in Calcalong Creek is lower than in typical KREEP (~20% of the usual abundance), indicating that it contains other components as well. This is to be expected as Calcalong Creek is a polymict microbreccia. For example, its composition lies in the highland field of Wood's⁹ diagram of Ti against Ca/Al diagram (Fig. 2*b*), yet it just meets the CaO/Al₂O₃ criterion established by Naney *et al.*¹⁰ for mare glasses, suggesting it contains a mare component. The compositions of other lunar meteorites shown to contain mare material, such as bulk samples of EET87521^{8,11}, lie in both highland and mare fields. Both Calcalong Creek and Y793274^{5,11} fall in the highland field and near the demarcation line defined by the polymict breccias 15558 and 79135. This is partly because all three meteorites are breccias in which each chip analysed may contain different proportions of highland and mare components.

The bulk composition of the new lunar meteorite seems to follow the trend that mare basaltic clasts and components in all lunar meteorites are of the low-titanium or very-low-titanium varieties^{5,11}. This is in contrast to the high-titanium mare basalts sampled and mapped by γ -ray spectrometry by the Apollo missions¹² and by ground-based reflectance spectroscopy¹³. Only a few of the basalts on the near side have been sampled¹³, however, and as we know that lunar meteorites are more representative samples of the moon, low- and very-low-titanium mare basalts may be more abundant than previously thought⁵. Mixing calculations indicate that Calcalong Creek is consistent with a mixture of 50% anorthosite, 20% KREEP, 15% Luna 16-type low-titanium mare basalt and 15% SCCRV (Sc-Cr-V) components. It contains the highest KREEP component of any lunar meteorite and is remarkably similar, in this respect, to samples collected at the Apollo 14 and 16 sites. □

TABLE 2 Important element weight ratios

	Mass	Fe/Mn	K/La	Fe/Sc	Ca/Al	References
Calcalong Creek 9,2,1	18 mg	73	91	3,220	1.03	this work
Calcalong Creek 9,2,2	6 mg	78	91	3,320	1.07	this work
Average Apollo soils		76	76	3,500	0.97	14
High-K KREEP		76	63	3,560	0.77	15, 16
Medium-K Fra Mauro		66	67	3,470	0.89	17
Low-K Fra Mauro		71	48	4,830	0.80	17
VHA basalt*		76	97	5,280	0.82	17
A17 VLT†		72		2,830	1.35	17
L16 LT‡		74	88	2,430	1.34	17
L24 VLT†		64	130	2,920	1.19	17
Anorthosite		—	95	2,360	0.72	17
Lunar meteorites						
Basaltic lunar meteorites						
EET87521		79	69	3,390	1.26	8
Y793274, 82		72	137	3,460	0.95	5
Anorthositic lunar meteorites (average)		70	110	3,700	0.81	18
Other meteorites						
angrites (average of 2)		99	5	94	—	8
brachinites (average of 2)		145	887	273	2.83	8
eucrites (average of 31)		35	122	5,040	—	8
Millbillillie (eucrite)	3 mg	36	100	5,400	1.50	this work
Diogenite		31	327	10,400	—	19
Howardite		34	126	6,200	—	19
SNC (average)		39	733	117	3.07	8

Element ratios are given for different Solar System samples for comparison with the Calcalong Creek samples.

* Very high aluminium.

† Very low titanium.

‡ Low titanium.

Received 16 January; accepted 25 June 1991.

- Score, R. & Lindstrom, M. M. *Antarct. Met. Newslett.* **13**, 84-132 (1990).
- Yanai, K. & Kojima, H. *Abstr. 53rd Met. Soc.* 187 (1990).
- Bevan, A. W. R. & Binns, R. A. *Meteoritics* **24**, 127-133 (1989).
- Delaney, J. S. *Nature* **342**, 889-890 (1989).
- Warren, P. H. & Kallemeyn, G. W. *Proc. NIPR Symp. Antarctic Meteorites* **4** (in the press).
- Stoffler, D. *et al.* *Proc. Conf. Lunar Highlands Crust* (eds Papke, J. J. & Merrill, R. B.) 51-70 (Pergamon, New York, 1980).
- Laul, J. C. & Schmitt, R. A. *Proc. 4th Lunar Sci. Conf.* (ed. Gose, W. A.) 1349-1367 (Pergamon, New York, 1973).
- Warren, P. H. & Kallemeyn, G. W. *Geochim. cosmochim. Acta* **53**, 3323-3300 (1989).
- Wood, J. A. *Proc. 6th Lunar Sci. Conf.* (ed. Merrill, R. B.) 1087-1102 (Pergamon, New York, 1975).

10. Naney, M. T. et al. *Proc. 7th Lunar Sci. Conf.* (ed. Merrill, R. B.) 155–184 (Pergamon, New York, 1976).
11. Lindstrom, M. M. et al. *Proc. NIPR Symp. Antarctic Meteorites 4* (in the press).
12. Metzger, A. E. et al. *Proc. 10th Lunar Planet. Sci. Conf.* (ed. Merrill, R. B.) 1719–1726 (Pergamon, New York, 1979).
13. Pieters, C. M. *Proc. 9th Lunar Planet. Sci. Conf.* (ed. Merrill, R. B.) 2825–2849 (Pergamon, New York, 1978).
14. Taylor, S. R. *Planetary Science: A Lunar Perspective* (Lunar and Planetary Institute, Houston, 1982).
15. Warren, P. H. & Wasson, J. T. *Rev. Geophys. Space Phys.* **17**, 73–88 (1979).
16. Warren, P. H. in *Workshop on Moon in Transition: Apollo 14 KREEP, and Evolved Lunar Rocks* (eds Taylor, J. G. Warren, P. H.) 149–152 (Lunar and Planetary Institute, Houston, 1989).
17. *Lunar Sourcebook, A User's Guide to the Moon* (eds Heiken, G. H. & Vaniman, D. T.) 736 (Cambridge University Press, Cambridge, 1991).
18. Palme, H. et al. *Abstr. 21st Lunar Planet. Sci. Conf.* XXI 930–931 (1990).
19. Fukuoka, T. et al. *Proc. 8th Lunar Sci. Conf.* (ed. Merrill, R. B.) 187–210 (Pergamon, New York, 1977).

The first surface faulting from a historical intraplate earthquake in North America

John Adams, R. J. Wetmiller, H. S. Hasegawa & J. Drysdale

Geophysics Division, Geological Survey of Canada,
1 Observatory Crescent, Ottawa K1A 0Y3; Canada

ALTHOUGH many fault ruptures are known from the boundaries of the Earth's lithospheric plates, the 'stable' interiors of continents are much less active¹. Worldwide, only ten historical intraplate earthquakes are known to have produced surface faulting² and none of these was in eastern North America. Despite the confirmed prehistoric rupture of the Meers fault³, some have wondered if intraplate earthquakes in the stable North American craton might somehow be different from plate-boundary earthquakes, and hence fail to rupture the surface⁴. This is an important consideration for seismic hazard assessments. When the Ungava earthquake (with magnitude (M_s) 6.3 and apparent shallow depth) occurred on 25 December 1989 in northern Canada, we (and, independently, A. Johnston of Memphis State University) concluded that it might have produced surface faulting. Here we report that there was indeed a surface rupture, 8.5 km long, with up to 1.8 m of reverse faulting on a steeply dipping, arcuate fault plane. The faulting confirms the seismological process and indicates that there is a hazard from ground-rupturing earthquakes in eastern North America, but the poor surface expression shows that finding prehistoric fault ruptures in order to constrain seismic hazard estimates will be difficult.

An immediate field study of the Ungava earthquake was precluded by its remoteness, December temperatures of -35°C , snow cover and short daylight (<3 h per day). We discovered the surface faulting during an aftershock survey in July 1990. The Lac Turquoise fault scarp (Fig. 1) lies north of the tree line in an area of continuous permafrost. The gneissic bedrock is typical of the heavily glaciated Superior Province of the Canadian Shield, which is 2.6 Gyr old. Local relief is only 70 m, and about half the region is covered by lakes, with Lac Bécarré surrounding the epicentre on three sides.

The trend of the rupture was evident from the torn and buckled muskeg, which had undergone brittle failure (it was frozen in December). Although the scarp seldom formed a free face, the offset could be measured in a few places (Fig. 2, arrows A and B), and in many places large differences in shoreline elevation occurred within a short distance (Fig. 1). In all cases the upthrown block was southeast of the fault.

The arcuate scarp is slightly concave to the northwest (average trend 038°), away from the upthrown block (Fig. 2). The sense of curvature in relation to the upthrown block is opposite to the curvature of the Australian surface faulting at Meckering⁵, Marryat Creek⁶ and Tennant Creek⁷. The Ungava rupture seems to follow the lake shorelines closely, sometimes deviating from a smooth concave arc by 25° . We infer that the faulting was



FIG. 1 Emerged, pre-earthquake shoreline deformed by the Lac Turquoise fault scarp at the northern end of Lac Turquoise, the first surface rupture by a historical intraplate earthquake in North America. The dark, stained band (arrows) that marks the old lake level is warped, and the light, unstained boulders below are visible only on the upthrown side of the fault (to the right). The boulders at the fault are about 1 m across. The 1.8-m throw on the fault at this point has been accentuated by partial drainage of the lake.

controlled by pre-existing bedrock structures which may also control the lake shorelines.

Toward the southwestern end of the rupture, there was systematic left-lateral offset of 0.1 m (four observations) on a short rupture segment striking at a high angle to the trend of the main reverse fault (Fig. 2). This was the only place we saw faulting cross the bedrock. Combined with a throw of 0.3 m on the reverse fault segment, the lateral offset yields a local dip of $\sim 70^\circ$ for the main fault.

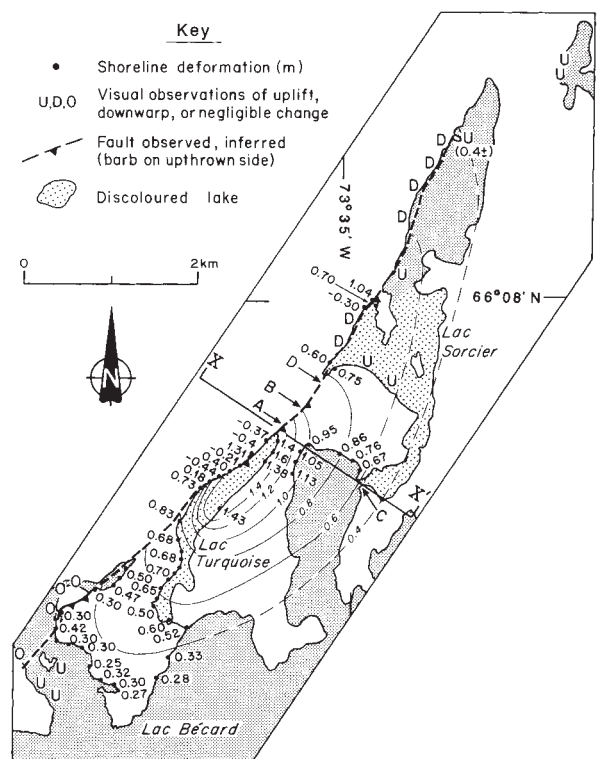


FIG. 2 Map of the surface rupture (centred at 60.12°N , 73.60°W), the two lakes still discoloured in July 1990, sites of shoreline measurement with corrected uplift relative to Lac Bécarré (some nearby sites lie on opposite sides of the fault), sites for which visual estimates were made from the air (U, D, O) and approximate contours of relative deformation inferred (dashed lines). On large boulders at lake level, shoreline change was measured directly to within a few centimetres. In most places, measurements were made by Abney level, and measurement errors were $\sim \pm 10\%$. Profile X-X' is shown on Fig. 4.

Deformed lake shorelines provided good evidence of the earthquake deformation. Uplifted shorelines were conspicuous because the newly-emerged rocks were black or reddish, which allowed a quick visual assessment of uplift, even from the air. The position and throw of the northern 2 km and southern 1 km of the 8.5-km-long rupture shown in Fig. 2 are based on clear evidence of uplifted shorelines observed from the air, as time permitted only four visits (totalling 17 h) on the ground.

The deformed region involved three lakes: Bécarré, Sorcier and Turquoise. Old air photographs show Lac Sorcier was sometimes a bay of Lac Bécarré (depending on seasonal water-level fluctuations), but 0.67 m of coseismic uplift of a strait (Fig. 2, arrow C) isolated the bay into a lake. In July, the water level in Lac Sorcier was 0.25 m higher than Lac Bécarré, and so 0.25 m should be added to the measurements of shoreline change made relative to Lac Sorcier. Because of earthquake-related underground drainage (see below), the level of Lac Turquoise was 1.8 m below its outlet into Lac Bécarré at the time of our visit. Near the outlet, adjacent differences in shoreline emergence on Lac Bécarré (0.5 m) and Lac Turquoise (1.8 m), together with extra uplift of 0.1 m for the outlet (because it is 200 m closer to the fault), yield a correction factor of ~ -1.2 m for Lac Turquoise observations made relative to the July water level. With these corrections of +0.25 m for Lac Sorcier and -1.2 m for Lac Turquoise observations to make them equivalent to the Lac Bécarré observations, the shoreline changes on all three lakes can be analysed together (Fig. 2). Lac Bécarré is large relative to the extent of the coseismic deformation; hence the contours closely approach the absolute elevation changes.

The deformation pattern is a simple elongate half-dome of uplift southeast of the fault and localized observations of subsidence to the northwest (Fig. 2). The maximum throw is restricted to the middle of the fault and tapers to <0.3 m at the ends (Fig. 3). Fault-dislocation modelling⁸ shows that the ratio of uplift to subsidence is very sensitive to fault dip and that the far-field uplift (not yet measured) is sensitive to fault width. For dips of $<80^\circ$, uplift dominates, as was observed. A 3-km-wide fault dipping at 70° provides an acceptable fit to the data (Fig. 4).

Sand volcanoes were found chiefly along the shores of Lac Turquoise, below the pre-earthquake lake level. We infer that continuous permafrost and seasonally frozen ground severely limited the amount of available water for liquefaction of the sediments. Many liquefaction-like features and evidence for groundwater outflow were found just south of Lac Sorcier (Fig. 2, arrow D). We hypothesize that Lac Turquoise drained underground along the fault rupture and that the silty water emerged to discolour Lac Sorcier.

The geological data indicate that the Ungava earthquake ruptured an arcuate, reverse-slip fault ~ 8.5 km long, and produced an average throw of 0.8 m. The fault dips steeply to the southeast and is unlikely to extend below 5 km. The average

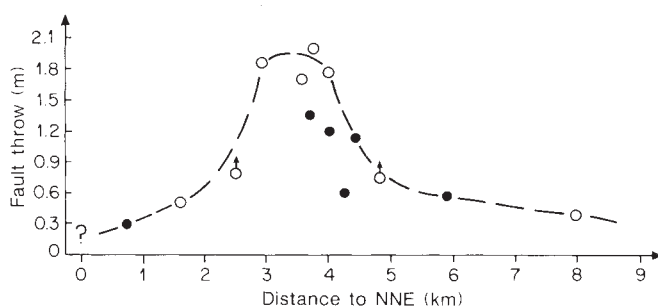


FIG. 3 Variation in throw along the Lac Turquoise fault scarp. The maximum uplift exceeds 1.4 m and the maximum subsidence is 0.55 m. Dashed line envelopes the data and suggests the trend. Data points with arrows indicate minimum estimates of throw. ●, Observation at or near fault; ○, from regional shoreline deformation.

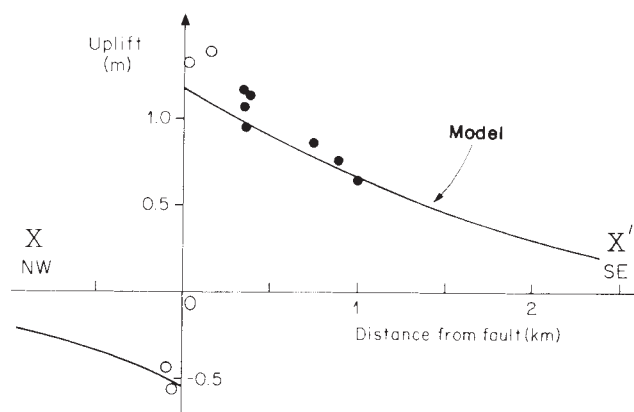


FIG. 4 Fault dislocation model and measurements along profile X-X' from Fig. 2. ●, represent shoreline uplift measured on Lac Bécarré, ○ are corrected uplift values on Lac Turquoise. The throw was observed at the fault, and the modelled fault dip and width (solid line) can be matched to more distant shoreline tilt. In the model, dip is 70° , width is 3 km and slip is 1.8 m.

throw, the fault length, a fault width of 3–5 km and a crustal rigidity of $3.3 \times 10^{10} \text{ N m}^{-2}$ give a seismic moment of $0.7\text{--}1.1 \times 10^{18} \text{ N m}$. The moment would be larger if there was an unrecognized strike-slip component (difficult to discern in the rough terrain) to the faulting. The published seismological moment⁹, $1.1 \times 10^{18} \text{ N m}$, is very similar to the geologically determined value, so, for the first time in eastern North America, the observed surface faulting confirms directly the amount of faulting inferred from distant seismograms.

Most previous intraplate North American earthquakes of similar size occurred at depths of ~ 10 km (ref. 10), and none produced surface faulting. The Ungava earthquake, being shallow and having a surface break, resembles the well-documented surface-rupturing earthquakes in Australia^{5–7} which, although in a similar cratonic environment, were hitherto considered anomalous. Thus, moderate earthquakes with similar large ground ruptures, which would be extremely damaging if in populated areas, seem possible on all continental shields. Together with the unusually deep (29 km) Saguenay earthquake of 1988 (ref. 11), the Ungava earthquake indicates that almost the entire thickness of continental shields and not just the deeper parts are capable of generating damaging earthquakes.

The historical record in eastern North America is short compared with the repeat time of the damaging earthquakes¹², so evidence of surface faulting left by prehistoric earthquakes would be useful to constrain seismic hazard estimates and complement studies of earthquake-induced liquefactions^{13,14} which provide less direct evidence. Unfortunately, our investigations indicate that if eastern North American earthquakes are similar to the Ungava earthquake (with its ephemeral poorly expressed fault scarps, torn muskeg and colour contrasts on lake-shore boulders), they will not leave a strong geological record of faulting. Hence, prehistoric surface ruptures elsewhere in the Canadian shield could easily be overlooked. Despite the indirect nature of the evidence, a silt layer in the lake sediments (for example, see ref. 15) may provide the best long-term record of the 1989 Ungava and perhaps of other large earthquakes on the craton. □

Received 18 March; accepted 30 May 1991.

1. Johnston, A. C. & Kanter, L. R. *Scient. Am.* **262**(3), 68–75 (1990).
2. Johnston, A. C. & Bullard, T. *Seismol. Res. Lett.* **61**, 152–153 (1990).
3. Crone, A. J. & Luza, K. V. *Bull. geol. Soc. Am.* **102**, 1–17 (1990).
4. Wu, F. T. *Seismol. Res. Lett.* **59**, 251–260 (1988).
5. Gordon, F. R. & Lewis, J. D. *Bull. geol. Surv. W. Austr.* **126**, 1–229 (1980).
6. McCue, K. et al. *EOS* **68**, 609; 612 (1987).
7. Choy, G. L. & Bowman, J. R. *J. geophys. Res.* **95**, 6867–6882 (1990).

8. Mansinha, L. & Smylie, D. E. *Bull. seismol. Soc. Am.* **61**, 1433-1440 (1971).
9. National Earthquake Information Service *Preliminary Determination of Epicenters*, December 1989 (Golden, Colorado, 1990).
10. Ebel, J. E., Somerville, P. G. & McIver, J. D. *J. geophys. Res.* **91**, 8231-8247 (1986).
11. North, R. G. et al. *Seismol. Res. Lett.* **60**, 89-93 (1989).
12. Basham, P. W. *Ann. N. Y. Acad. Sci.* **558**, 1-10 (1989).
13. Amick, D. & Gelenas, R. *Science* **251**, 655-658 (1991).
14. Obermeier, S. F. et al. *Science* **251**, 1061-1063 (1991).
15. Doig, R. *Can. J. Earth Sci.* **23**, 930-937 (1986).

ACKNOWLEDGEMENTS. We thank A. C. Johnston, T. Bullard, A. Crone and our colleagues for their help.

Mating patterns in seminatural populations of mice influenced by MHC genotype

Wayne K. Potts*, C. Jo Manning*†
& Edward K. Wakeland*

* Laboratory of Molecular Genetics, Department of Pathology and Laboratory Medicine, University of Florida, Gainesville, Florida 32610, USA
† Department of Psychology, University of Washington, Seattle, Washington 98195, USA

BECAUSE of the central role of major histocompatibility complex (MHC) genes in immune recognition¹⁻³, it is often assumed that parasite-driven selection maintains the unprecedented genetic diversity of these genes⁴⁻⁷. But associations between MHC genotype and specific infectious diseases have been difficult to identify^{8,9} with a few exceptions such as Marek's disease¹⁰ and malaria¹¹. Alternatively, MHC-related reproductive mechanisms such as selective abortion¹²⁻¹⁵ and mating preferences^{16,17} could be responsible for the diversity. To determine both the nature and strength of selection operating on MHC genes by we have studied components of selection in seminatural populations of mice (*Mus musculus domesticus*). Here we assess MHC-related patterns of reproduction and early (preweaning) mortality by analysing 1,139 progeny born in nine populations, and 662 progeny from laboratory matings. Reproductive mechanisms, primarily mating preferences, result in 27% fewer MHC-homozygous offspring than expected from random mating. MHC genotype had no detectable influence on neonatal (preweaning) mortality. These mating preferences are strong enough to account for most of the MHC genetic diversity found in natural populations of *Mus*.

We⁶ and others^{18,19} have argued that the artificiality of using inbred strains under laboratory conditions makes it impossible to use such methodologies to assess accurately the evolutionary forces operating on MHC genes in the complexity of natural populations. Our approach to solving these problems was to study populations of outbred *Mus* in enclosures of sufficient size and complexity to allow normal patterns of social competition and reproductive behaviour under normal pathogen loads. Mice were recently derived from wild-caught individuals, but carried MHC haplotypes (b, d, k and q) from four inbred strains (see Table 1 for breeding details). Founding populations consisted of 24-31 age-matched and MHC-genotyped individuals. Unique ear punch patterns allowed individual identification (with binoculars) permitting documentation of social status, territorial boundaries, home ranges, mating pairs and other social behaviours. Pups were trapped near weaning age (15-21 days), earmarked, and 2-cm tail biopsies taken for DNA samples. After about 150 progeny were born, each population was disbanded and both preweaning stage offspring and embryos from pregnant females were MHC genotyped.

Table 1 summarizes the observed and expected frequencies for MHC haplotypes, heterozygotes and homozygotes. Although there were some departures from expected haplotypic frequencies, the mean change for the b, d, k and q haplotypes was small (0.027, 0.033, -0.026 and -0.036, respectively), indicating that

no haplotype showed systematic gains or losses over the nine populations.

The only strong and consistent departure from expected frequencies was the deficiency of MHC homozygous progeny shown by all populations, with a mean deficiency of 27% (Table 1). The combined χ^2 for these homozygote deficiencies was highly significant ($P < 0.0001$). As a large proportion of the 1,139 progeny were genotyped near the age of weaning, the deficiency of homozygotes could have resulted from differential neonatal mortality. This hypothesis was excluded because the proportion of homozygotes did not decline between the embryo and nestling stages (Table 2). These results indicate that the observed deficiency of homozygotes must be due to mechanisms operating before birth.

To test for MHC-related selective fertilization or abortion, we conducted informative laboratory matings. Although there was a 6% deficiency of MHC homozygotes relative to expected mendelian ratios, this difference was not statistically significant ($P = 0.16$; Table 2). Further, if the genotypic data in Table 1 are adjusted for this potential 6% contribution, the remaining deficiency of MHC homozygotes (mean = 21%) is still significant ($P < 0.001$).

The only remaining explanation for the homozygote deficiencies is MHC-based nonrandom mating. To determine if these mating preferences operated through nonrandom settlement patterns associated with MHC genotypes, females were assigned to the male in whose territory her sleeping/nest box was located. When these settlement pairs were used to recalculate the expected homozygosities in Table 1 (originally calculated assuming random mating between all females and all territorial males), the mean deficiency of homozygotes decreased from 27% to 19%, suggesting that settlement patterns do account for a portion of the homozygote deficiency.

The largest portion of the homozygote deficiency is still unaccounted for and must be due to extraterritorial matings, which

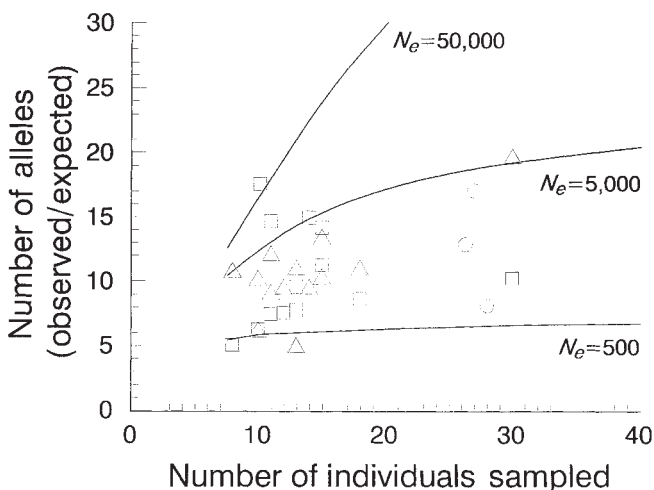


FIG. 1 The number of observed MHC alleles from local *Mus* populations (symbols) plotted with the expected number of alleles (lines) for three effective population sizes (N_e), under levels of disassortative mating that would yield a 27% deficiency of homozygotes. The expected values (corrected for sample size) were derived from an analytical model for symmetric overdominant selection³⁰, using a selection coefficient of $s = 0.27$ and a mutation rate of 10^{-6} (an empirically derived estimate for MHC loci³¹). As N_e for *Mus* is uncertain, the expected number of alleles across the entire estimated range of N_e values (500-50,000) is provided⁶. Data for class I K ($\square$) and D ($\triangle$) loci are from serological analysis of 13 European populations³². The total number of alleles was estimated by assuming that the allelic frequency distributions of blanks and identified alleles were the same. Data for the class II A_β locus ($\circ$) are from a restriction fragment length polymorphism study of three North American populations⁶.

TABLE 1 Observed and expected frequencies in progeny of MHC haplotypes, heterozygotes and homozygotes

Population	Number of founders ♂ ♀		Haplotypic frequency of offspring								MHC heterozygotes obs. exp.		MHC homozygotes obs. exp.		Homozygote deficiency (%)	P
			b		d		k		q							
			obs.	exp.	obs.	exp.	obs.	exp.	obs.	exp.						
			A	8	16	0.34	0.31	0.41	0.31	0.12	0.30	0.13	0.09	93	87.4	25
B	8	16	0.26	0.27	0.10	0.06	0.28	0.36	0.36	0.32	88	82.9	24	29.1	18	0.27
C	8	16	0.26	0.15	0.44	0.39	0.04	0.10	0.26	0.37	102	96.7	39	44.3	12	0.39
D	15	15	0.32	0.23	0.46	0.44	0.12	0.08	0.11	0.26	89	82.8	30	36.2	17	0.22
E	7	16	0.29	0.32	0.12	0.12	0.24	0.20	0.34	0.37	121	107.1	24	37.9	37	0.009
F	12	19	0.15	0.10	0.54	0.60	0.04	0.04	0.28	0.26	104	83.1	40	60.9	34	0.0004
G	8	16	0.23	0.21	0.32	0.29	0.13	0.21	0.32	0.29	131	121.6	26	35.4	27	0.07
H	8	16	0.29	0.27	0.31	0.26	0.20	0.18	0.20	0.29	99	90.1	20	28.9	31	0.06
I	8	16	0.07	0.12	0.34	0.26	0.26	0.22	0.34	0.41	72	58.1	12	21.6	44	0.006

Expected genotypic frequencies were calculated as $x_i y_i + x_j y_j$, where x and y are the frequencies in females and territorial males, respectively, of the i th and j th haplotypes. Nonterritorial males were excluded from the calculation of all expected values because this provided a better predictor of allelic frequencies of progeny which is consistent with our behavioural and genetic data indicating nonterritorial males do not breed. Excluding nonterritorial males also controlled for the only differential reproductive success found to be correlated with MHC, the ability of males to gain territories (manuscript in preparation). Regression analysis: when departures from expected haplotypic frequencies were regressed on expected haplotypic frequencies, a significant inverse correlation was found (slope = -0.94 , $P < 0.025$; $r^2 = 0.11$, $P < 0.01$), indicating that the frequencies of common haplotypes are decreasing whereas rare haplotypes are increasing. This pattern is predicted by many forms of diversity-maintaining selection. **Methods. Mice.** Animals used in these experiments came from generations three and six of original crosses between wild-caught animals and four inbred strains (C57BL/10J, BALB/c, B10.BR and DBA/1, carrying MHC haplotypes b, d, k and q, respectively). In the F2-generation, only mice homozygous for one of the four inbred-derived MHC haplotypes were used to continue this outbred colony. In the resulting outbred animals, half of the genome is wild-derived and the other half is derived from one or more inbred strains. The number of inbred strains constituting the inbred-derived half of the genome was manipulated in conjunction with experiments designed to test the relative importance of inbreeding at the remainder of the genome (manuscript in preparation). In populations A–C, MHC homozygotes were derived from a single inbred strain, whereas MHC heterozygotes were derived from two inbred strains. This established a systematic correlation between inbreeding at MHC loci and genome-wide inbreeding, a condition that would be expected in natural populations that experience some inbreeding²⁶, as is the case for *Mus*. In our breeding design this correlation involved 1/4 of the genome and only loci derived from inbred strains. In populations G–I this correlation was eliminated by crosses that systematically introduced uniform contributions of all four inbred strains in each individual. Populations D–F were founded by progeny born in the enclosure to the founders of populations A–C, respectively. Consequently, populations D–F differed from the other six populations in two important respects. First, they were born and reared in a natural social system, in contrast to being cage-reared. Second, some of the male–female combinations were sibs, a condition that was avoided in populations A–C and G–I. To control for the possibility that behavioural mechanisms exist to inhibit matings between sibs, the random mating expectations for populations D–F were calculated with potential sib matings excluded. **MHC genotyping.** Restriction fragment length polymorphisms from *TaqI*-digested genomic DNA (extracted from 2 cm tail biopsies) allowed identification of all 10 MHC genotypes. Southern blots were hybridized with a 5.8-kilobase *EcoRI* fragment cloned from a *Mus* MHC class II A_β gene. Protocols for DNA extraction, Southern blotting and hybridization are detailed elsewhere²⁷. **Enclosures.** The outdoor (but predator free) enclosures (4.9 m $\times$ 9.8 m) were subdivided with 46-cm-high screening (0.5-cm grids) into eight equal subsections, each containing an additional 3 m spiral of screening. The screening created spatial complexity thought to be important for normal behaviour²⁸. Mice could easily climb the screening, but it was often used as a territorial boundary. Observation periods of 1–2 h at dusk and daytime nest checks were made 5–7 times per week. **Pathogen load.** A variety of factors suggest that the enclosure populations were subjected to relatively normal pathogen loads. We frequently introduce wild-caught mice to our colony, providing a continuous supply of endemic mouse pathogens. The enclosure is surrounded by pasture land allowing normal vectoring of pathogens found in commensal populations of mice. When 55 mice from population C were tested for the presence of antibodies to the endemic pathogen *Mycoplasma pulmonis*, all individuals tested positive.

seem to be controlled by females. We observed 41 matings. All occurred in the territory of the mating male and 13 (32%) of these were extrapair matings, where the female had travelled to a nearby territory. Males were never observed following oestrous females beyond their own territorial boundaries. Paternity analysis of litters where maternity was unambiguous (58 litters, 305 pups), revealed that 52% of litters involved extraterritorial matings (a minimum estimate as genetic exclusions were based solely on the four MHC haplotypes). In the litters involving extraterritorial sires, there were 41% fewer MHC homozygous offspring than expected if the female had mated solely with her territorial male ($P = 0.001$; Table 3), suggesting that females

seek extraterritorial matings with males that are relatively more MHC-disparate than their own territorial male.

The precise nature of MHC-based mating preferences awaits further experimentation, but it is generally assumed that they are disassortative¹⁶. Some forms of disassortative mating have population dynamics similar to overdominance²⁰, which allows the use of well developed overdominant models to estimate the importance of these mating preferences in maintaining MHC genetic diversity in *Mus*. Figure 1 compares the number of MHC alleles observed in natural populations with the number expected from mating preferences of the magnitude observed in this study. Such preferences can account for most of the MHC polymorphism in natural populations^{6,7}.

The above results suggest that MHC-based mating preferences are a major force responsible for the maintenance of MHC

TABLE 2 Observed and expected frequencies of MHC heterozygotes and homozygotes for nestlings and embryos (enclosure populations) and from laboratory matings

Enclosure populations	Heterozygotes		Homozygotes		Homozygote deficiency (%)
	obs.	exp.	obs.	exp.	
Nestlings	682	619	186	247	25
Embryos	217	194	54	77	30
Laboratory matings					
♀ ♂					
a/a $\times$ a/b	90	85	80	85	6
a/b $\times$ a/a	92	81.5	71	81.5	13
a/b $\times$ a/b	125	122.5	120	122.5	2
a/b $\times$ b/c	63	63	21	21	0
Totals/means	370	352	292	310	6

For the two laboratory mating classes with expected homozygote proportions of 1/4 (a/b $\times$ a/c) and 1/2 (a/b $\times$ a/b and a/a $\times$ a/b), no significant differences in fecundity were found either in mean litter size (6.3 and 6.4, respectively; $P > 0.4$) or in days to first litter (25 and 28, respectively; $P > 0.2$). These results indicate that fecundity was not correlated with the degree of MHC disparity between mates.

TABLE 3 Extraterritorial matings result in more MHC heterozygous offspring than expected if the female had mated with her territorial male

Type of mating	Heterozygotes		Homozygotes		Homozygote deficiency (%)	P*
	obs.	exp.	obs.	exp.		
Within territory matings	114	113.5	28	28.5	2	0.92
Extra-territorial matings	137	118.6	26	44.4	41	0.001

Expected values were calculated on the basis of females mating with the male in whose territory her nest/sleeping box was located. These data are derived from the subset of litters where maternity could be unambiguously assigned. Determination of maternity through observation was often impossible owing to the communal nesting and nursing behaviour of *Mus*²⁹.

* χ^2 test.

genetic diversity in *Mus*. The laboratory demonstration that mice²¹, rats²² and humans²³ can, through olfaction, discriminate between MHC congenic strains of rodents, provides a mechanistic basis for these MHC-based mating preferences. Two alternative, but not mutually exclusive, functions have been hypothesized for the evolution of MHC-based mating preferences. Progeny from these matings may have increased fitness because either their MHC genotype confers increased disease resistance (for example through heterozygote advantage)^{6,16}, or MHC serves as a polymorphic marker in a genetic incompatibility system that functions to minimize genome-wide inbreeding^{6,18,24,25}. □

Received 22 April; accepted 20 June 1991.

1. Babbitt, B. P., Allen, P. M. & Matsuoka, G. *Nature* **317**, 359–361 (1985).
2. Roy, S., Scherer, M. T., Briner, T. J., Smith, J. A. & Geffer, M. L. *Science* **244**, 572–575 (1989).
3. Kappler, J. W., Roehm, N. & Marrack, P. *Cell* **49**, 273–280 (1987).
4. Doherty, P. C. & Zinkernagel, R. M. *Nature* **256**, 50–52 (1975).
5. Hughes, A. L. & Nei, M. *Nature* **335**, 167–170 (1988).
6. Potts, W. K. & Wakeland, E. K. *Trends Ecol. Evol.* **5**, 181–187 (1990).
7. Takahata, N. & Nei, M. *Genetics* **124**, 967–978 (1990).
8. Tiwari, J. L. & Terasaki, P. I. *HLA and Disease Associations* (Springer, New York, 1985).
9. Klein, J. *Natural History of the Major Histocompatibility Complex* (Wiley, New York, 1986).
10. Pazderka, F., Longenecker, B. M., Law, G. R. J., Stone, H. A. & Ruth, R. F. *Immunogenetics* **2**, 93–100 (1975).

11. Hill, A. V. S. *et al. Nature* (this issue).
12. Clarke, B. & Kirby, D. R. S. *Nature* **211**, 999–1000 (1966).
13. Palm, J. *Cancer Res.* **34**, 2061–2065 (1974).
14. Hamilton, M. & Hellstrom, I. *J. Reprod. Immun.* **19**, 267–270 (1978).
15. Thomas, M. L., Harger, J. H., Wagner, D. K., Rabin, B. S. & Gill, T. J. *Am. J. Obstet. Gynec.* **151**, 1053–1058 (1985).
16. Yamazaki, K. *et al. J. exp. Med.* **144**, 1324–1335 (1976).
17. Egid, K. & Brown, J. L. *Anim. Behav.* **38**, 548–549 (1989).
18. Partridge, L. *Phil. Trans. R. Soc. B* **319**, 525–539 (1988).
19. Nei, M. & Hughes, A. L. in *Evolution at the Molecular Level* (eds Selander, R. K., Clark, A. G. & Whittam, T. S.) 222–247 (Sinauer, Sunderland, Massachusetts, 1991).
20. Karlin, S. & Feldman, M. W. *Genetics* **59**, 117–136 (1968).
21. Yamazaki, K. *et al. J. exp. Med.* **150**, 755–760 (1979).
22. Singh, P. B., Brown, R. E. & Roser, B. *Nature* **327**, 161–164 (1987).
23. Gilbert, A. N., Yamazaki, K. & Beauchamp, G. K. *J. comp. Pathol.* **100**, 262–265 (1986).
24. Brown, J. L. in *Ethical Questions in Brain and Behaviour* (ed. Pfaff, D. W.) 111–124 (Springer, New York, 1983).
25. Uyenoyma, M. K. in *The Evolution of Sex* (eds Michod, R. E. & Levin, B. R.) 212–232 (Sinauer, Sunderland, Massachusetts, 1988).
26. Weir, B. S. & Cockerham, C. C. *Genet. Res.* **21**, 247–262 (1973).
27. McConnell, T. J., Talbot, W. S., McIndoe, R. A. & Wakeland, E. K. *Nature* **332**, 651–654 (1988).
28. Bronson, F. H. Q. *Rev. Biol.* **54**, 265–299 (1979).
29. Wilkinson, G. S. & Baker, A. E. M. *Ethology* **77**, 103–114 (1988).
30. Yokoyama, S. & Nei, M. *Genetics* **91**, 609–626 (1979).
31. Melvold, R. W. & Kohn, H. I. in *Transgenic Mice and Mutants in MHC Research* (eds Egorov, I. K. & David, C. S.) (Springer, New York, 1990).
32. Nadeau, J. H., Wakeland, E. K., Gotze, D. & Klein, J. *Genet. Res.* **37**, 17–31 (1981).

ACKNOWLEDGEMENTS. We thank M. Brown for *Mycoplasma* screening, M. Uyenoyma for comments on the manuscript, and technicians and students who contributed to the collection of behavioural and molecular data. The work was supported by the NIH.

Development and function of T cells in mice rendered interleukin-2 deficient by gene targeting

Hubert Schorle, Thomas Holtschke, Thomas Hünig, Anneliese Schimpl & Ivan Horak*

Institute of Virology and Immunobiology, University of Würzburg, Versbacherstrasse 7, 8700 Würzburg, Germany

INTERLEUKIN-2 (IL-2) is a lymphocytotropic hormone which is thought to have a key role in the immune response of mammalian cells. It is produced by a subpopulation of activated T-lymphocytes and acts *in vitro* as the principal auto- and paracrine T-cell growth factor (for reviews see refs 1–3). IL-2 is, however, not the sole T-cell growth factor^{4,5}, nor does it act exclusively on T cells, also promoting growth of NK cells⁶ and differentiation of B cells⁷. A role for IL-2 in T-cell development has been postulated but remains controversial^{8–12}. Here we test the requirement for IL-2 *in vivo* using IL-2-deficient mice generated by targeted recombination. We find that mice homozygous for the IL-2 gene mutation are normal with regard to thymocyte and peripheral T-cell subset composition, but that a dysregulation of the immune system is manifested by reduced polyclonal *in vitro* T-cell responses and by dramatic changes in the isotype levels of serum immunoglobulins.

To inactivate the endogenous IL-2 gene by targeted recombination^{13–16}, a neomycin-resistance (*neo^r*) gene was inserted in reverse orientation into the third exon of the IL-2 genomic clone¹⁷ (Fig. 1). This insertion introduces several stop codons in all reading frames which abolish the biological activity of the IL-2 protein completely¹⁸. The mutated fragment was introduced into embryonic stem cells (ES cells)¹⁹ and *neo^r* colonies were selected. Clones with the correct homologous recombination event at the IL-2 locus were identified by polymerase chain reaction (PCR) and Southern blot analysis. The frequency of the correctly targeted allele was one in 200.

Two ES cell clones carrying the disrupted IL-2 gene on one allele were injected into blastocysts of C57BL/6 mice and

transferred to CD-1 foster mothers. Eight out of 40 animals born were chimaeras. To test for germ-line transmission of the IL-2 mutation, the chimaeras were mated with C57BL/6 mice. Three out of five chimaeric mice derived from the same clone transmitted the ES cell genome to their progeny. Mice heterozygous for the disrupted IL-2 gene were mated and their progeny analysed to detect animals homozygous for the mutation (IL-2^{-/-}) (Fig. 1d). As expected, the IL-2 mutation was transmitted in a mendelian fashion.

Intrathymic differentiation of T cells expressing the T-cell antigen receptor (TCR) proceeds from a TCR-negative, CD4⁻ through a TCR-low, CD4⁺ stage towards the mature TCR-high, CD4⁺ and CD4⁺ phenotypes that seed the peripheral lymphoid organs (reviewed in ref. 20). Both positive selection (for self MHC-restriction) and negative selection (for self tolerance) occur during passage through the CD4⁺ compartment²¹. To compare the representation of these thymocyte subsets in normal and in IL-2-deficient animals, a litter of seven mice derived from IL-2^(+/-) parents was analysed by two-colour flow cytometry at 4 weeks of age (Fig. 2). It is apparent that disruption of the IL-2 gene had no major effect on thymocyte subset composition. Both immature and mature subsets were present at normal levels in animals homozygous for the inactivated gene. Also, the small contribution of γ/δ T cells was unaffected (not shown). Finally, the average total cell number per thymus was similar in normal and mutant mice (2.20×10^8 ($n = 7$) versus 2.18×10^8 ($n = 5$); data from two litters). These findings rule out an essential role for the IL-2/IL-2R-pathway in thymopoiesis and are in full agreement with a reported case of human IL-2 deficiency²².

Cell-surface marker analyses of spleen (not shown) and lymph node cells (Fig. 2) also revealed no striking differences between normal and IL-2-deficient mice. These observations clearly indicate that IL-2 is neither essential for thymocyte maturation nor for seeding of the periphery with mature T cells. IL-2 deficiency was, however, manifested at the functional level. Cells from thymus, lymph node and spleen of IL-2-deficient mice responded less well than those from normal littermates to polyclonal T-cell activators (concanavalin A, Table 1; and anti-CD3, data not shown) unless IL-2 was added. Because, as expected, no IL-2 activity was detected in supernatants of concanavalin A-activated spleen and lymph node cells from IL-2-deficient mice the residual responses after polyclonal activation suggest the use of alternative growth factor pathways.

* To whom correspondence should be addressed.

TABLE 1 Proliferation and IL-2 production in response to conA

Genotype	Culture conditions†	Thymus		³ H thymidine incorporation (c.p.m. × 10 ⁻³)*		Lymph node		CTLL assay‡
		-IL-2	+IL-2	-IL-2	+IL-2	-IL-2	+IL-2	
+/+	Medium	0.06	3.92	3.45	20.61	2.50	59.60	31.62
	conA	34.60	197.46	37.82	118.13	66.95	172.85	
+/+	Medium	0.09	5.23	3.39	9.93	0.37	9.30	21.12
	conA	42.00	187.64	37.73	97.15	18.45	58.68	
+/-	Medium	0.06	6.12	3.92	26.24	1.77	19.05	27.44
	conA	24.40	183.24	39.00	84.64	45.74	125.71	
+/-	Medium	0.01	5.27	1.86	76.96	1.04	25.63	28.56
	conA	68.80	206.73	49.58	128.75	46.66	130.08	
-/-	Medium	0.20	6.54	2.49	51.17	1.67	23.67	0.70
	conA	18.60	215.27	25.79	180.06	18.35	178.64	
-/-	Medium	0.13	19.23	1.12	28.46	2.67	27.43	0.39
	conA	6.10	205.55	6.51	55.33	17.46	144.38	
-/-	Medium	0.05	5.84	2.30	59.14	2.88	75.06	0.56
	conA	11.12	191.78	3.24	113.33	10.49	147.38	

Thymocytes, spleen cells or lymph node cells of seven littermates from an IL-2 +/+ mating were stimulated with concanavalin A (conA) in the absence or presence of 20 U ml⁻¹ human recombinant IL-2 and assayed for proliferation and IL-2 production.

* Means of triplicates from cultures pulsed with 0.25 µCi (2 Ci mM⁻¹) from 48 to 64 h of culture (T-cell activation) or 24–48 h (IL-2 indicator line CTLL-2).

† Thymocytes (5 × 10⁵ per well), spleen or lymph node cells (2 × 10⁵ per well) were cultured in 0.2 ml RPMI/5% FCS in 96-well flat-bottom trays. ConA: 2.5 µg ml⁻¹.

‡ 24-h supernatants of conA-activated spleen cells (2 × 10⁶ ml⁻¹) were assayed at 1:4 final dilution on 5 × 10³ CTLL-2 cells per microwell. Background (medium only): 3,490 c.p.m.

Another functional deficiency was revealed when short-term CD4⁺ T-cell lines derived from the lymph nodes of normal and mutated animals were investigated for their ability to help anti-µF(ab')₂-primed B cells to secrete IgM. Figure 3 shows that the T cells from the minus mutants were totally unable to overcome the negative effect on differentiation which is observed in

proliferating B cells after massive B cell receptor engagement²³. Addition of exogenous IL-2 could overcome the deficit. To test whether the lack of IL-2 formation had any influence on the isotype distribution and concentration in nonimmunized mutant mice, the sera from the litter described above were analysed at 4 weeks of age. IgM levels of mutant and normal mice were

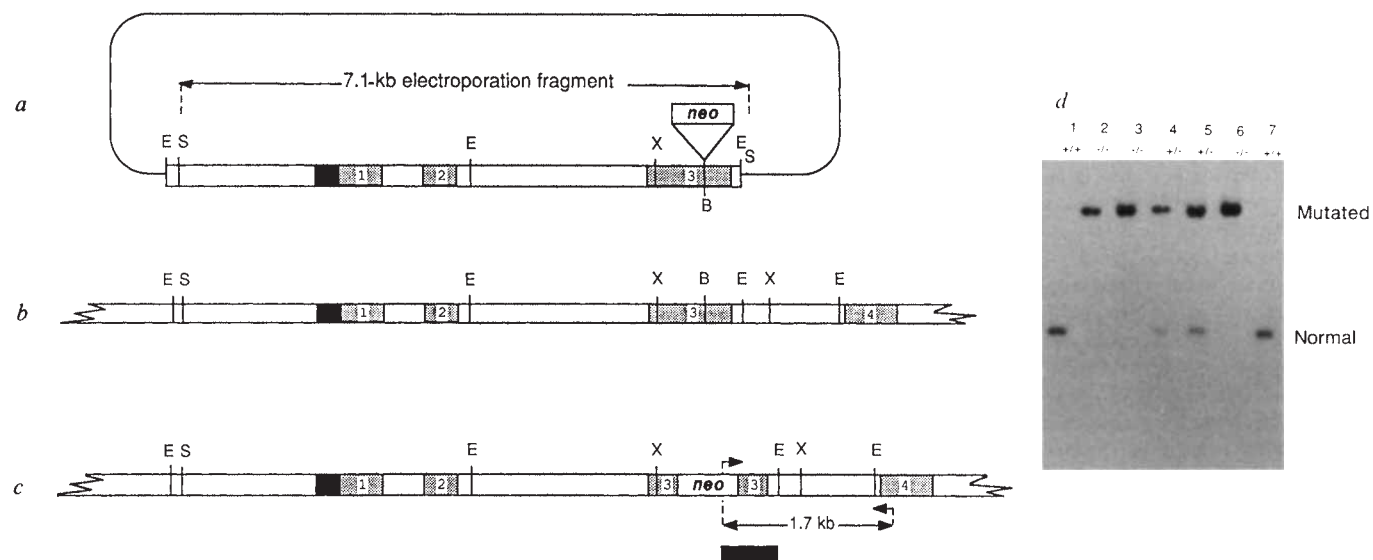


FIG. 1 Targeted disruption of the IL-2 gene. *a*, The *neo*^r cassette of pMC1neopA (ref. 14) was inserted in reverse orientation into the *Bgl*III site of the mouse IL-2 gene cloned in pUC 19. This 1.1-kilobase (kb) insertion interrupted the reading frame of the IL-2 gene in its third exon. The arrows indicate the 7.1-kb *Sma*I fragment which was used for electroporations. *b*, A scheme of an intact mouse IL-2 allele. *c*, The expected mutation of the IL-2 allele resulting from homologous recombination with the mutated IL-2 gene fragment depicted in *a*. The arrow heads mark the positions of primers used for PCR. The 1.7-kb fragment indicated is the resulting product of the PCR. The Southern probe, a 600-base pair *Eco*RI fragment containing *neo*^r and exon 3 sequences is indicated by a black bar. Regulatory sequences of the IL-2 gene are indicated by black boxes, exons are shadowed. B, *Bgl*III; E, *Eco*RI; S, *Sma*I; X, *Xba*I. *d*, Analyses of IL-2 locus in offspring of heterozygous mice (IL-2^{+/-} × IL-2^{+/-}). Tail DNA was digested with *Xba*I and analysed by Southern blotting. The *neo*^r/exon 3 hybridization probe is described in

Fig. 1a. Animals 1 and 7 are IL-2^{+/+}; 4 and 5 are IL-2^{+/-}; 2, 3 and 6 are IL-2^{-/-}.

METHODS. ES cells (cell line E14, derived from 129/Ola mice (ref. 25)) were grown in DMEM (high glucose, Gibco) with 10% FCS, 0.1 mM 2-mercaptoethanol, nonessential amino acids, and 10³ U ml⁻¹ of leukaemia inhibitory factor^{26,27} (LIF, Amrad). Cells were grown as bulk cultures on mitomycin C-treated, mouse embryo-derived fibroblasts in the absence of antibiotics. The ES cells were maintained without feeder cells for three passages before electroporation. ES cells (1 × 10⁶) were electroporated with 30 µg DNA and grown in LIF-supplemented media without feeder cells. After 3 days, the cells were divided. The replica plates designed for the PCR test were kept for 6 days with 250 µg ml⁻¹ G418 (Gibco), followed by 3 days without the drug. The replica plates of ES cells for blastocyst injections were selected for 11 days with G418.

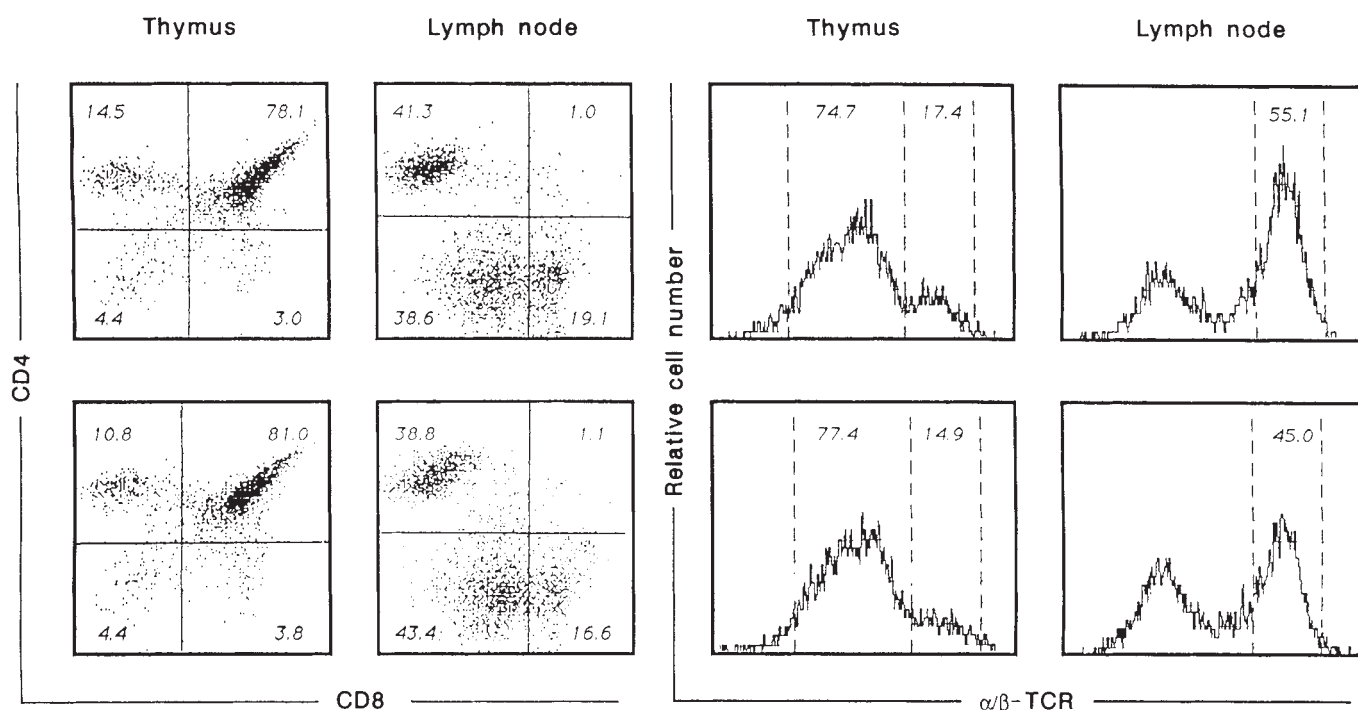


FIG. 2 Expression of CD4, CD8 and the α/β -TCR in IL-2-deficient mice. Thymocytes and lymph node cells from a normal and a mutant littermate were analysed by indirect immunofluorescence and flow cytometry. METHODS. Single-cell suspensions of thymus or pooled lymph nodes (10^5 cells) were stained with anti-CD8 monoclonal antibody followed by FITC-coupled rabbit-anti-rat immunoglobulin (Dakopatts GmbH, Hamburg), 10% normal rat serum and phycoerythrin-coupled monoclonal antibody GK1.5 to CD4 (Becton Dickinson, Mountain View, California). Alternatively,

they were stained with biotinylated antibody 57-597-21 to the α/β -TCR (ref. 28) and streptavidin-phycoerythrin (Becton Dickinson). All treatments were in PBS, 0.2% BSA, 0.1% NaN₃ for 15 min on ice. Cells were analysed using a FACscan (Becton Dickinson) flow cytometer with scatter gates set to include all viable cells and logarithmic amplification of fluorescence signals. Two thousand cells were analysed using the LYSYS software (Becton Dickinson) and are displayed as correlated dot plots or histograms.

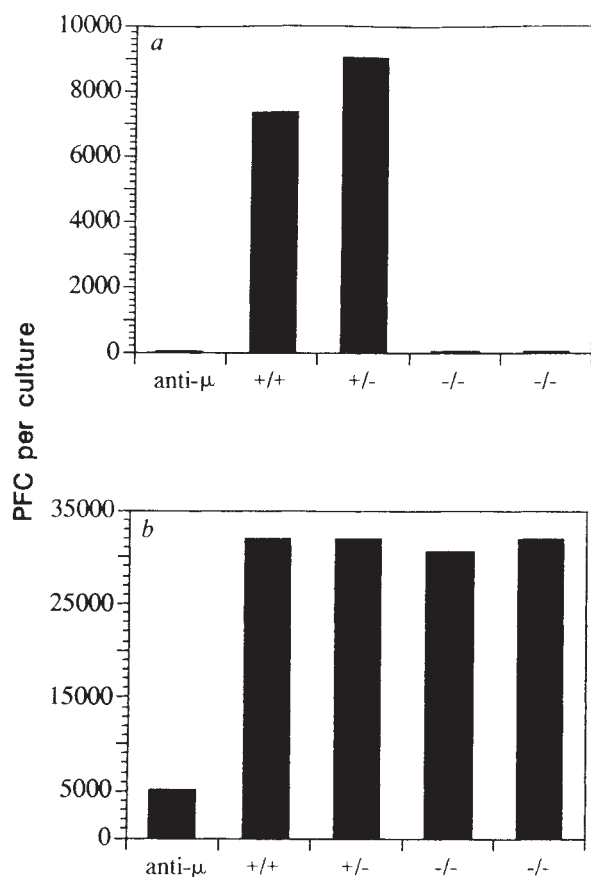


FIG. 3 Inability of T cells from IL-2-deficient mice to induce IgM secretion in anti μ F(ab')₂-stimulated B cells. Resting B cells (1×10^5 ; Percoll density 1.08) isolated from normal B6D2F1 spleen were stimulated with 5 μ g anti μ F(ab')₂ (Jackson, Dianova, Hamburg). T cells (5×10^4 per culture) were added at day 0 without (a) or with (b) exogenous mouse recombinant IL-2 (5 U ml^{-1}). IgM secreting cells were analysed 5 days later in a polyclonal plaque assay²⁹. PFC, plaque-forming cells.

comparable ($320\text{--}620 \mu\text{g ml}^{-1}$ versus $530\text{--}620 \mu\text{g ml}^{-1}$). By contrast, dramatic differences were observed in the serum concentrations of IgG1 ($1,600\text{--}2,700 \mu\text{g ml}^{-1}$ versus $<50\text{--}60 \mu\text{g ml}^{-1}$). IgG2a and 2b were also higher (data not shown).

The immune system of IL-2 deficient mice has not yet been put to the test of combating microbial infections. Only such a challenge may reveal the necessity for IL-2 expression *in vivo*. An increased susceptibility to infection was indeed found in human immunodeficiency due to an IL-2 defect^{22,24}. Thus in addition to being an excellent tool to evaluate pleiotropic effects of IL-2 in the complex regulatory circuits of the immune system and the role of IL-2 in innate and adaptive immunity to pathogens, the IL-2^(-/-) mice also provide a model for the treatment of a human immunodeficiency disease caused by a defective lymphokine gene. □

Received 2 April; accepted 1 July 1991.

- Smith, K. A. *Science* **240**, 1169–1176 (1988).
- Paul, W. E. *Cell* **57**, 521–524 (1989).
- Mosmann, T. R. *et al. Prog. Immun.* **7**, 611–618 (1989).
- Lee, F. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 2061–2065 (1986).
- Chazen, G. D., Pereira, G. M. B., LeGros, G., Gillis, S. & Shevach, E. M. *Proc. Natn. Acad. Sci. U.S.A.* **86**, 5923–5927 (1989).

6. Trinchieri, G. *et al.* *J. exp. Med.* **160**, 1147–1169 (1984).
7. Tigges, M. A., Casey, L. S. & Koshland, M. E. *Science* **243**, 781–786 (1989).
8. Takacs, L., Osawa, H. & Diamantstein, T. *Eur. J. Immun.* **14**, 1152–1156 (1984).
9. Jenkinson, E. J., Kingston, R. & Owen, J. J. T. *Nature* **329**, 160–162 (1987).
10. Tentori, L., Longo, D. L., Zuniga-Pflucker, J. C., Wing, C. & Kruisbeek, A. M. *J. exp. Med.* **168**, 1741–1747 (1988).
11. Plum, J. & de Smedt, M. *Eur. J. Immun.* **18**, 795–799 (1988).
12. MacDonald, H. R. *et al.* *Immunol. Rev.* **104**, 157–182 (1988).
13. Doetschman, T. *et al.* *Nature* **330**, 576–578 (1987).
14. Thomas, K. R. & Capechi, M. R. *Cell* **51**, 503–512 (1987).
15. Thompson, S., Clarke, A. R., Pow, A. M., Hooper, M. L. & Melton, D. W. *Cell* **56**, 313–321 (1989).
16. Zijlstra, M., Li, E., Sajjadi, F., Subramani, S. & Jaenisch, R. *Nature* **342**, 435–438 (1989).
17. Fuse, A. *et al.* *Nucleic Acids Res.* **12**, 9323–9331 (1984).
18. Zurawski, S. M. & Zurawski, G. *EMBO J.* **7**, 1061–1069 (1988).
19. Evans, M. J. & Kaufman, M. H. *Nature* **292**, 154–156 (1981).
20. Shortman, K., Egerton, M., Spangrude, G. J. & Scollay, R. *Semin. Immun.* **2**, 3–12 (1990).
21. von Boehmer, H. & Kieselow, P. *Science* **248**, 1369–1372 (1990).
22. Weinberg, K. & Parkman, R. *New Engl. J. Med.* **322**, 1718–1723 (1990).
23. Chen-Bettecken, U., Wecker, E. & Schimpl, A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7384–7388 (1985).
24. Pahwa, R. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **86**, 5069–5073 (1989).
25. Hooper, M. *et al.* *Nature* **326**, 292–295 (1987).
26. Smith, A. G. & Hooper, M. L. *Dev. Biol.* **121**, 1–9 (1987).
27. Williams, R. L. *et al.* *Nature* **336**, 684–687 (1988).
28. Kubo, R., Born, W., Kappler, J., Marrack, P. & Pigeon, M. *J. Immun.* **142**, 2736–2742 (1989).
29. Molinaro, G. A., Maron, E., Eby, W. C. & Dray, S. *Eur. J. Immun.* **5**, 771–774 (1975).

ACKNOWLEDGEMENTS. We thank M. Hooper for ES cells, R. Swoboda for IL-2 DNA clones, H. Haber, R. Mitnacht and K. Borschert for assistance, and S. Siddell for critical reading of the manuscript. These studies were supported by the Deutsche Forschungsgemeinschaft and by the Fonds der Chemischen Industrie.

Making antibody fragments using phage display libraries

Tim Clackson*, Hennie R. Hoogenboom†, Andrew D. Griffiths† & Greg Winter*†

*MRC Laboratory of Molecular Biology and †Centre for Protein Engineering, Hills Road, Cambridge CB2 2QH, UK

To by-pass hybridoma technology and animal immunization, we are trying to build antibodies in bacteria by mimicking features of immune selection¹. Recently we used fd phage² to display antibody fragments fused to a minor coat protein^{3,4}, allowing enrichment of phage with antigen³. Using a random combinatorial library of the rearranged heavy (VH) and kappa (VK) light chains^{5–8} from mice immune to the hapten 2-phenyloxazol-5-one (phOx), we have now displayed diverse libraries of antibody fragments on the surface of fd phage. After a single pass over a hapten affinity column, fd phage with a range of phOx binding activities were detected, at least one with high affinity (dissociation constant, $K_d = 10^{-8}$ M). A second pass enriched for the strong binders at

the expense of the weak. The binders were encoded by V genes similar to those found in anti-phOx hybridomas but in promiscuous combinations (where the same V gene is found with several different partners). By combining a promiscuous VH or VK gene with diverse repertoires of partners to create hierarchical libraries, we elicited many more pairings with strong binding activities. Phage display offers new ways of making antibodies from V-gene libraries, altering V-domain pairings and selecting for antibodies with good affinities.

We used the polymerase chain reaction (PCR)⁹ to amplify the VH and VK genes from the spleen messenger RNA of mice immunized with phOx, and also developed a 'PCR assembly' process¹⁰ to link these genes together randomly for expression as single-chain Fv (scFv) fragments^{11,12} (Fig. 1a–c). The assembled genes were cloned in a single step into the vector fdDOG1 (Fig. 1e) for display as a fusion with the fd gene III coat protein. This initial library of 2×10^5 clones seemed to be diverse (Fig. 1d), and sequencing revealed the presence of most VH groups¹³ and VK subgroups¹⁴ (data not shown). None of the 568 clones tested bound to phOx as detected by enzyme-linked immunosorbent assay (ELISA).

The library of phages was passed down a phOx affinity column (Table 1a), and eluted with hapten. Of the eluted clones, 13%

TABLE 1 Affinity selection of hapten-binding phage

	Precolumn	Clones binding to phOx*		
		After first round	After second round	After third round
(a) Random combinatorial libraries				
phOx-immunized mice	0/568 (0%)	48/376 (13%)	175/188 (93%)	—
Unimmunized mice	—	—	0/388 (0%)	—
(b) Hierarchical libraries				
VH-B/VK-rep library	6/190 (3%)	348/380 (92%)	—	—
VH-rep/VK-d library	0/190 (0%)	23/380 (7%)	—	—
(c) Fractionation of VH-B/VK-d and VH-B/VK-b phage*				
Mixture of clones	88/1,896 (4.6%) (44/1,740 (2.5%)+)	55/95 (57.9%)	1,152/1,156 (99.7%)	1,296/1,299 (99.8%)

Selection of phage with hapten-binding activities from the random combinatorial and hierarchical libraries (a and b, respectively), and fractionation of clones with different affinities for phOx (c). For the random combinatorial libraries fdDOG1 RF was extensively digested with *NotI* and *ApaI*, purified by electroelution²⁴ and 1 µg ligated to 0.5 µg (5 µg for the hierarchical libraries) of the assembled scFv genes in 1 ml with 8,000 units T4 DNA ligase (New England Biolabs) overnight at 16 °C. Purified ligation mix was electroporated in six aliquots into MC1061 cells²⁵ and plated on NZY medium²⁴ with 15 µg ml⁻¹ tetracycline, in 243 × 243 mm dishes (Nunc); 90–95% of clones contained scFv genes by PCR screening (see legend to Fig. 1). Colonies were scraped into 50 ml 2 × TY medium²⁶ and shaken at 37 °C for 30 min. Liberated phage were precipitated twice with polyethylene glycol and resuspended to 10¹² transducing units (TU) ml⁻¹ in water (titred as in ref. 3). For affinity selection, a 1-ml column of phOx-BSA-Sepharose²⁷ M. Dreher and C. Milstein, unpublished results) was washed with 300 ml PBS, and 20 ml PBS containing 2% skimmed milk powder (MPBS). Phage (10¹² TU) were loaded in 10 ml MPBS, washed with 10 ml MPBS and finally 200 ml PBS. The bound phage were eluted with 5 ml 1 mM 4-ε-amino-caproic acid methylene 2-phenyl-oxazol-5-one (phOx-CAP). About 10⁶ TU eluted phage were amplified by infecting 1 ml log phase *E. coli* TG1 (ref. 28) and plating as above. For a further round of selection, colonies were scraped into 10 ml 2 × TY medium and then processed as above. For the hierarchical libraries, VH-B and VK-d genes were individually recombined, then assembled with the VH or VK repertoires. For the fractionation of clone VH-B/VK-d, 7 × 10¹⁰ TU phage in the ratio 20 VH-B/VK-b:1 VH-B/VK-d were loaded onto a phOx-BSA-Sepharose column in 10 ml MPBS and eluted as above. Eluted phage were used to reinfect *E. coli* TG1, and phage produced and harvested as before. About 10¹¹ TU of phage were loaded onto a second affinity column and the process repeated to give a total of three column passes. Dilutions of eluted phage at each stage were plated in duplicate and probed separately²⁴ with oligonucleotides specific for VK-b (5'-GAGCGGGTAACCACTGTACT) or VH-d (5'-GAATGGTATAGTACTACCCT).

* In (c), numbers refer to VH-B/VK-d.

† Numbers after three reinfections and cycles of growth. This control, omitting the column steps, confirms that a spurious growth or infectivity advantage was not responsible for the enrichment of clone VH-B/VK-d.

bound to phOx, and ranged from poor to strong binding in ELISA. We sequenced 23 of these hapten-binding clones and found eight different VH genes (A-H) in a variety of pairings with seven different V κ genes (a-g) (Fig. 2a). Most of the domains, such as VH-B and V κ -d, were able to bind hapten with any of several partners¹⁵. The probability of finding multiple partners for a given chain should depend mainly on the inherent promiscuity of the chain and on the number of available partners and competing chains. Two other examples of promiscuous pairings have been noted in random combinatorial libraries made in λ phage^{6,8}, so this may prove to be a feature of small combinatorial libraries from immunized animals.

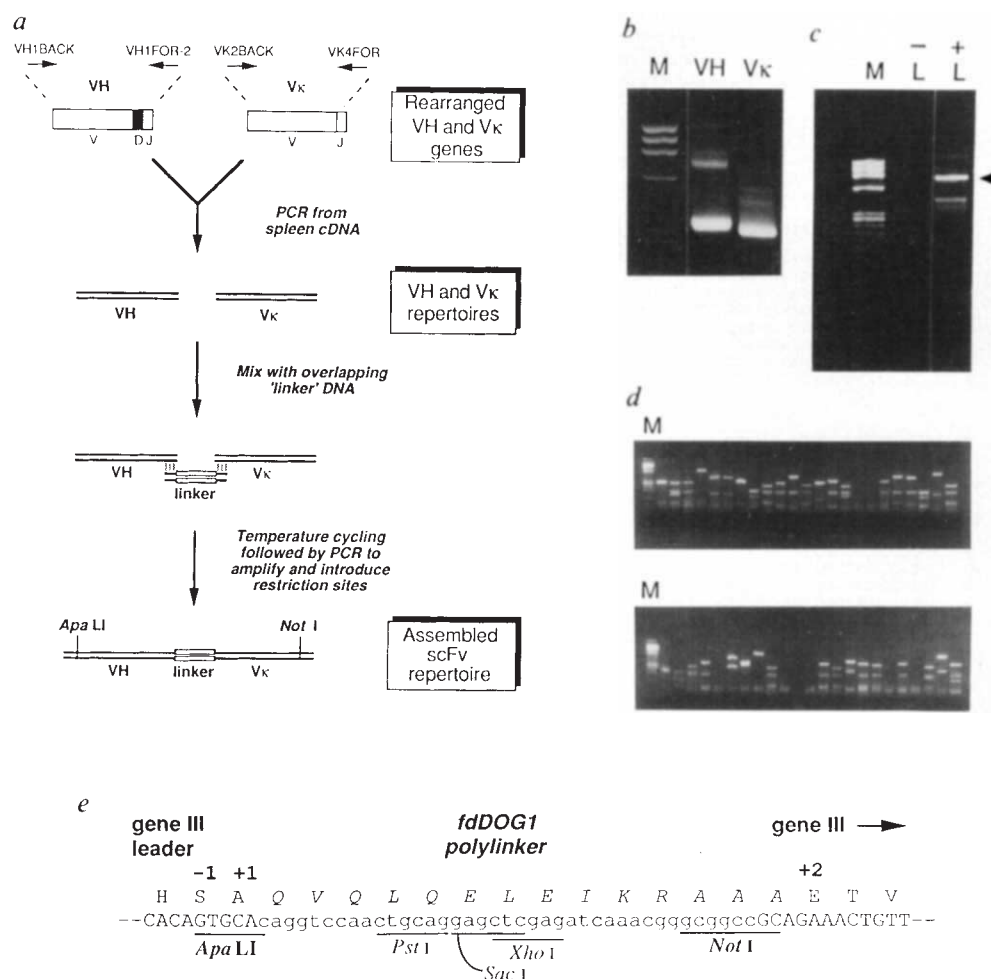
The sequences of the V genes were related to those seen in the secondary response to phOx, but with differences (Fig. 2b). Thus most phOx hybridomas from the secondary response use somatically mutated derivatives of three types of V κ genes, V κ ox1, 'V κ ox-like' and V κ 45.1 genes¹⁶. These can pair with VH genes from several groups, but V κ ox1 more commonly pairs with the VHox1 gene (VH group 2; ref. 13). V κ ox1 genes are

always, and V κ ox-like genes often, found in association with heavy chain genes (including VHox1) that encode a short five-residue CDR3, with the sequence motif Asp-X-Gly-X-X (where X is any amino acid¹⁶), in which the central glycine creates a cavity for phOx (ref. 17). In our library, nearly all of the VH genes belonged to group 1, and most of the V κ genes were ox-like and associated with VH genes encoding a five-residue CDR3 motif Asp/Asn-X-Gly-X-X (Fig. 2b). V κ ox1 and VHox1 were found only once (V κ -f and VH-E), and not in combination with each other: indeed V κ -f does not encode the Trp 91 involved in phOx binding¹⁷ and was paired with a VH gene (VH-C) encoding a six-residue CDR3.

The promiscuity of the VH-B and V κ -d domains prompted us to force further pairings, by assembling these genes with the entire repertoires of either V κ or VH genes from the same immunized mice. The resulting hierarchical libraries, (VH-B $\times$ V κ -rep and VH-rep $\times$ V κ -d), each with 4×10^7 members, were subjected to a round of selection and hapten-binding clones isolated (Table 1b). Most were strong binders by ELISA

FIG. 1 PCR assembly of scFv library. **a**, VH and V κ genes are separately amplified, then mixed with a linker fragment that overlaps them both. The linker (93 base pairs) encodes the short peptide, (Gly₄Ser)₃, which links VH and V κ in scFvs (ref. 11). Cycles of annealing-denaturation, followed by reamplification of the mixture, generate a random combinatorial cassette of VH and V κ genes joined in-frame for expression. **b**, VH and V κ gene repertoire PCR products from the immunized mice analysed by electrophoresis on agarose (1%) gel. **c**, PCR assembly of scFv gene repertoires with linker (+L) or without (-L); arrow indicates assembled repertoire. M is DNA marker Φ X174 replicative form DNA digested with *Hae*III. **d**, Diversity of library as seen by *Bst*NI fingerprinting of individual clones. **e**, Sequence of fd gene III around the signal peptide cleavage site in fdDOG1.

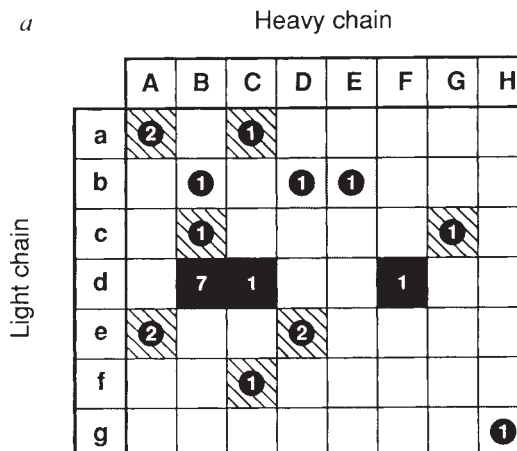
METHODS. For the random combinatorial libraries, cytoplasmic RNA was isolated²⁹ from the pooled spleens of either 5 male BALB/c mice boosted 8 weeks after primary immunization with phOx coupled to chicken serum albumin²⁷, or of 5 unimmunized mice. The cDNA was made with avian myeloblastosis virus reverse transcriptase (Anglian Biotech)³⁰ and primers that straddle the junction between the variable and constant regions of γ heavy chains and κ light chains (C. Marks, unpublished data). VH and V κ repertoires were amplified from the cDNA with 25 cycles of PCR (94 °C for 1 min, 60 °C for 1 min, 72 °C for 2 min) using Vent polymerase (New England Biolabs) and the primers VH1BACK (ref. 19) and VH1FOR-2 (ref. 31) or the primers VK2BACK and VK4FOR. The linker DNA was similarly amplified from pSW2scD1.3 (ref. 3) using primers LINKFOR and LINKBACK (complementary to VK2BACK and VH1FOR-2 respectively). After gel purification, 1 μ g each of the VH and V κ products were mixed with 300 ng linker in a 25 μ l PCR reaction mix without primers, and cycled 7 times (94 °C 2 min, 72 °C 4 min) to join the fragments, then amplified for 20 cycles (94 °C 1.5 min, 72 °C 2.5 min) using 25 pmol each VH1BACK and VK4FOR primers. Finally, the assembled products were gel-purified and reamplified with VH1BACK-ApaLI and VK4FOR-NotI ('tagged' versions of the original primers) to append restriction sites. Products (1–5 μ g) were extensively digested with ApaLI



and NotI for cloning into fdDOG1. Recombinant colonies were screened by PCR³² with the primers VH1BACK and VK4FOR, followed by digestion with the frequently cutting enzyme *Bst*NI. Primers: VK2BACK 5'-GACATTGAGCTC-ACCCAGTCTCCA; VK4FOR, an equimolar mix of 5'-CCGTTTGATTTCAGCTT-GGTGCC, 5'-CCGTTTATTTCCAGCTTGGTCCC, 5'-CCGTTTATTTCCAACCTTTG-TCCC and 5'-CCGTTTCAGCTCCAGCTTGGTCCC; LINKFOR, 5'-TGGAGACTGGGT-GAGCTCAATGTC; LINKBACK, 5'-GGGACCACGGTCACCGTCTCCTCA; VH1BACK-ApaLI, as VH1BACK (ref. 19) but with 5'-CATGACCACAGTGCAC added at the 5' end; VK4FOR-NotI, as VK4FOR but with 5'-GAGTCATTCTCGGCCCGC similarly added (restriction sites underlined).

FIG. 2 *a*, Matrix of V_H and V_K genes identified in phOx-binding clones selected from random combinatorial library. The number of clones found with each combination is shown. The binding to phOx-BSA, as judged by the ELISA signal, seemed to vary (marked by shading): no binding was seen to BSA alone. Optical density at 405 nm: 0.2–0.9, dotted box; 0.9–2.0, hatched box; >2.0, solid box. *b*, Encoded protein sequences of phOx-binding clones. Sequences of phOx-binding clones isolated (single-letter amino-acid code) after one round of selection of the random combinatorial library, with pairings as above, or the hierarchical library. Note that the first eight or seven residues, and the last nine or eleven residues, of the V_K or V_H genes, respectively, are encoded by the PCR primers. Classifications into V_H groups¹³ and V_K subgroups¹⁴, and the position of residue 91 encoded by the V_K genes (#), are indicated. The relationship to genes from the hybridoma analysed secondary response to phOx (ref. 16) is also shown; all of the V_K genes are 'ox-like', apart from those marked * with an asterisk, which are V_Kox1, and the only example of V_H group 2 (V_H-E) is V_Hox1. The intensity of ELISA signals from the hierarchical libraries, corrected relative to the signal from control phage, are indicated: Optical density at 405 nm 0.9–2.0 (+), >2.0 (+++). Multiple isolations of sequences are marked, and sequences (V_H-B and V_K-c) isolated from the random combinatorial library, and also the hierarchical libraries, are shown in *italics*. The V_H-B/V_K-d and V_H-B/V_K-c pairings gave similar signals (after correction of ELISA) when recovered from either combinatorial or hierarchical libraries.

METHODS. We screened for binding of the phage to hapten by ELISA: 96-well plates were coated with 10 µg ml⁻¹ phOx-BSA²⁷ or 10 µg ml⁻¹ BSA in PBS overnight at room temperature. Colonies of phage-transduced bacteria were inoculated into 200 µl 2 × TY medium²⁶ with 12.5 µg ml⁻¹ tetracycline in 96-well plates ('cell wells', Nuclon) and grown with shaking (300 r.p.m.) for



24 h at 37 °C. At this stage, cultures were saturated and phage titres were reproducible (10¹⁰ TU ml⁻¹). Phage supernatant (50 µl), mixed with 50 µl PBS containing 4% skimmed milk powder, was then added to the coated plates. Further details given in ref. 3. To sequence the clones, template was prepared²⁴ from the supernatants of 10-ml cultures grown for 24 h, and sequenced using the dideoxy method³³ and a Sequenase kit (USB), with primer LINKFOR for the V_H genes and primer fdSEQ1 (5'-GAATTTCTGTATGAGG) for the V_K genes.

b

V_H sequences

From combinatorial library:

From combinatorial library:				CDR1	CDR2	CDR3	Group	ELISA signal	
A	QVQLQSGAELARPGASVKMSCKASGYTFT	SYTMH	VVKRPPGGGLEWIG	YINPSSGYTNYNQKFKD	KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR	RYGAY	WGQGTTVTVSS	x4	1
B	QVQLQSGAELARPGASVKMSCKASGYTFT	RDWMH	VVKRPPGGGLEWIG	YINPSTGYTEYNQKFKD	KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR	NYGLY	WGQGTTVTVSS	x9	1
C	QVQLQSGGPELVKPGASVKMSCKASGYTFT	SYVMH	VVKRPPGGGLEWIG	YINPNDGTNYNEKFKG	KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR	YRFFPY	WGQGTTVTVSS	x3	1
D	QVQLQSGGPELVKPGASVKISCKASGYSTFT	GYFMH	VVKSHGSKLEWIG	RINPNDGTNYNQKFKG	KATLTVDKSSSTAHMELSLTSEDSAVYYCVG	ITTRFAY	WGQGTTVTVSS	x3	1 (see Fig. 2a)
E	QVQLQESGGLVAPQSLSITCTVSGFSLT	SYGVH	VVKRPPGGGLEWIG	VWAGGSTNYNSALMS	RLSISKDNKSKQVFLKMNLSLQDDTAMYYCAR	DRGDY	WGQGTTVTVSS	2	1
F	QVQLQSGAELARPGASVKMSCKASGYTFT	SYLMH	VVKRPPGGGLEWIG	YINPSTGYTEYNQKFKD	KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR	DYGYI	WGQGTTVTVSS	1	1
G	QVQLQSGAELARPGASVKLSCKASGYTFT	RYLMH	VVKRPPGGGLEWIG	YINPSTGYTEYNQKFKD	EATLTADKSSSTAYMQLSSLTSEDSAVYYCAR	DYGYI	WGQGTTVTVSS	1	1
H	QVLOQSGGPELVKPGASVKISCKASGYSTFS	RNYMH	VVKSHGSKLEWIG	YIAPFNGGTYNYQKFKG	KATLTVDKSSSTAYMHLSSLTSEDSAVYYCAT	DVGDR	WGQGTTVTVSS	1	1

From hierarchical library V_H-rep × V_K-d:

I	QVQLQSGPELARPGVSKMSCKASGYTFT	SYAMH	VWKQSQSKSLEWIG	VISTYNGNTNYNQKFKG	KATMTVDKSSSTAYMELARLTSEDSAIYYCAR	DYGDY	WGQGTTVTVSS	1	+++
J	QVQLQSGAELARPGASVKMSCKASGYTFT	RYTMH	VWKQRPGGLEWIG	YINPSSGYTNYNQKFKD	KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR	DRGAY	WGQGTTVTVSS	1	+++
K	QVQLQSGAELARPGASVKMSCKASGYTFT	RDWMH	VWKQRPGGLEWIG	YINPSTGYTEYNQKFKD	KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR	NYGLY	WGQGTTVTVSS	x3	1
L	QVQLQSGLELARPGASVKMSCKASGYTFT	NYLMH	VWKQRPGGLEWIG	YINPSTGYTEYNQKFKD	KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR	DYGYI	WGQGTTVTVSS	x2	1
M	QVQLQSGAELARPGASVKMSCKASGYTFT	NYVMH	VWKQRPGGLEWIG	YINPSTGYTEYNQKFKD	KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR	DYGYI	WGQGTTVTVSS	1	+++
N	QVQLQSGAELVKPGASVKLSCKTSGYFT	SYTMH	VWKQRPGGLEWIG	YINPSSGYTNYNQKFKD	KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR	DYGYI	WGQGTTVTVSS	1	++
O	QVQLQSGAELARPGASVKMSCEASGYTFT	SHLMH	VWKQRPGGLEWIG	YINPRTGYTEYNQKFKD	KATFTADKSSSTAYMQLSSLTSEDSAVYYCAR	DYGYI	WGQGTTVTVSS	1	++
P	QVQLQSGAELARPGASVKMSCKASGYTFT	SYVMH	VWKQRPGGLEWIG	YINPSTGYTEYNQKFKD	KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR	DYGYI	WGQGTTVTVSS	1	++
Q	QVQLQSGAELARPGASVKMSCKATGYTFT	SYLMH	VWKQRPGGLEWIG	YINPSTGYTEYNQKFKD	KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR	DYGYI	WGQGTTVTVSS	1	+++
R	QVQLQSGAELARPGASVKMSCKASGYTFT	SYVMH	VWKQRPGGLEWIG	YINPSSGYTNYNQKFKD	KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR	NYGYI	WGQGTTVTVSS	1	++
S	QVQLQSGAELARPGASVKMSCKASGYTFT	TFLMH	WLKQRPGGLEWIG	YINPSTGYTEYNQKFKD	KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR	DYGYI	WGQGTTVTVSS	x2	1
T	QVQLQSGAELARPGASVKMSCKASGYTFT	SYTMH	VWKQRPGPGGLEWIG	YINPSSGYTNYNQKFKD	KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR	DYGYI	WGQGTTVTVSS	x6	1
U	QVQLQSGAELARPGASVKMSCKASGYTFT	SYTMH	VWKQRPGGLEWIG	YINPPTGYTEYNQKFKD	KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR	DYGYI	WGQGTTVTVSS	1	+++
V	QVQLQSGAELARPGASVKMSCKASGYTFT	RDWMH	WLKQRPGGLEWIG	YINPSTGYTEYNQKFKD	KATLTADKSSSTAYMQLSSLTSEDSAVYYCAR	NYGLY	WGQGTTVTVSS	1	+++

V_K sequences

From combinatorial library:

From combinatorial library.									
	CDR1			CDR2			CDR3		
a	DIELTQSPSSLSASLGERVSLTIC	RASQISGYLS	WLQKPPGDSIKRLIY	AASTLES	GVPKRFSGSGSGTYSYSLTISSEAEADAA	YQYASYP	FGAGTKLEIKRA	x3	V
b	DIELTQSPAIMSASPGEKVTITC	RASSSVSSYLH	WYQKSGASPKRWIY	STSNLAS	GVPARFSGSGSGTYSYSLTISSEAEADAA	YQYSGYPT	FGAGTKLEIKRA	x3	IV
c	DIELTQSPITMAASPGEKVTITC	RASSSVSSYLH	WYQKPGSPKLLIY	RTSNLAS	GVPARFSGSGSGTYSYSLTIGTMAEADVA	YQSSITPT	FGAGTKLEIKRA	x2	IV (see
d	DIELTQSPITMAASPGEKVTITC	RASSSVSSYLH	WYQKPGSPKLLIY	RTSNLAS	GVPARFSGSGSGTYSYSLTIGTMAEADVA	YQSSITPT	FGAGTKLEIKRA	x9	IV Flg. 2a)
e	DIELTQSPAIMSASPGEKVTITC	RASSSVSSYLH	WYQKPGTSPKLLIY	STSNLAS	GVPTRFSGSGSGTYSYSLTISRMAEADAA	YQSSYPPT	FGSGTKLEIKRA	x4	VI
f	DIELTQSPAIMSASPGEKVTITC	RASSSVSSYLH	WYQKSGTSPKRWIY	DTSKLAS	GVPARFSGSGSGTYSYSLTISSEAEADAA	YQSSNPPT	FGAGTKLEIKRA	x1	VI *
g	DIELTQSPAIMSASPGEKVTITC	RASSSVSSYLH	WYQKSGTSPKRWIY	DTSKLAS	GVPARFSGSGSGTYSYSLTISSEAEADAA	YQSSNPPT	FGGGTKLEIKRA	VI	

From hierarchical library V_H-B × V_K-rep:

h	DIELTQSPAIMSASPGEKVTITC	SASSSVSYMH	WYQKSGTSPKRWIY	DTSKLAS	GVPARFSGSGSGTYSYSLTISSMEEADAATYYC	QQWSSNPILT	FGAGTKLEIKRA	x4	VI	*	+++
i	DIELTQSPAIMSASPGEKVTITC	SASSSVSYIH	WYQKSGTSPKLIWI	STSNLAS	GVPARFSGSGSGTYSYSLTISRMEEADAATYYC	QQYHSYPLT	FGAGTKLEIKRA	VI	++		
j	DIELTQSPITMAASPGKIIITC	SASSISSNYLH	WYQKPGFSPKLIWI	RTSNLAS	GVPARFSGSGSGTYSYSLTIGTMEAEADVATYYC	QQGSSPIPLT	FGGGTKLEIKRA	IV	++		
k	DIELTQSPITMAASPGDMITITC	SATSSISSNYLH	WYQKPGFSPKLIWI	RTSNLAS	GVPPRFSGSGSGTYSYSLTIGAMEAEADVATYYC	QQGSSPIPYT	FGAGTKLEIKRA	IV	++		
l	DIELTQSPITMAASPGKIIITC	SASSISSNYLH	WYQKPGFSPKLIWI	RTSNLAS	GVPARFSGSGSGTYSYSLTIGTMEAEADVATYYC	QQGSSPIPYT	FGGGTKLEIKRA	IV	++		
m	DIELTQSPITMAASPGKIIITC	SASSISSNNLH	WYQKPGFSPKLIWI	RTSNLAS	GVPARFSGSGSGTYSYSLTIGTMEAEADVATYYC	QQGSSPIPYT	FGGGTKLEIKRA	IV	++		
n	DIELTQSPITMAASPGKIIITC	SASSISSNYLH	WYQKPGFSPKLIWI	RTSNLAS	GVPARFSGSGSGTYSYSLTIGTMEAEADVATYYC	QQGSSIPFT	FGGGTKLEIKRA	IV	++		
o	DIELTQSPITMAASPGKIIITC	SASSISSNYLH	WYQKPGFSPKLIWI	RTSNLAS	GVPARFSGSGSGTYSYSLTIGTMEAEADVATYYC	QQGSSPIPYT	FGGGTKLEIKRA	x2	IV	++	
p	DIELTQSPAIMSASPGEKVTITC	SASSSVSYMH	WYQKSGTSPKRWIY	DTSKLAS	GVPARFSGSGSGTYSYSLTISSMEEADAATYYC	QQWSSNPILT	FGAGTKLEIKRA	x2	VI	*	+++
q	DIELTQSPAIMSASPGEKVTITC	SASSSVRYVN	WYQKSGTSPKRWIY	DTSKLAS	GVPARFSGSGSGTYSYSLTISSMEEADAATYYC	QQWTSNPNT	FGGGTKLEIKRA	VI	++		
r	DIELTQSPAIMSASPGEKVTITC	SASSSVSYMH	WYQKSGTSPKRWIY	DTSKLAS	GVPARFSGSGSGTYSYSLTISSMEEADAATYYC	QQWSTNALT	FGAGTKLEIKRA	VI	*	+++	
s	DIELTQSPAIMSASPGEKVTITC	RASSSVTSSYLH	WYQKSGASPKLWVY	STSNLAS	GVPARFSGSGSGTYSYSLTISSEVEADAATYYC	QQYSGYPLT	FGAGTKLEIKRA	IV	++		
t	DIELTQSPAIMSASPGEKVTITC	RASSSVSSSYLH	WYQKSGASPKLIWI	STSNLAS	GVPARFSGSGSGTYSYSLTISRMEEADAATYYC	QQRSYSYPLT	FGAGTKLEIKRA	IV	++		
u	DIELTQSPAIMSASPGEKVTITC	RASSSVSSYLH	WYQKSGASPKLIWI	STSNLAS	GVPARFSGSGSGTYSYSLTISSEVEADAATYYC	QQYSGYPLT	FGAGTKLEIKRA	IV	++		
v	DIELTQSPAIMSASPGEKVTITC	RASSSVSSSYLH	WYQKSGASPKLIWI	STSNLPS	GVPARFSGSGSGTYSYSLTISSEVEADAATYYC	QQYSGYPLT	FGGGTKLEIKRA	IV	++		
w	DIELTQSPITMAASPGKIIITC	SASSISSNYLH	WYQKPGFSPKLIWI	RTSNLAS	GVPARFSGSGSGTYSYSLTIGTMEAEADVATYYC	QQGSSPIPLT	FGAGTKLEIKRA	x3	IV	++	

(Fig. 2b). By sequencing 23 clones from each library, we identified 14 new partners for VH-B and 13 for V κ -d; apart from VH-B and V κ -c, none of the previous VH-B or V κ -d partners (or indeed other partners) cloned and sequenced from the random combinatorial library was isolated again. These features are consistent with the much larger number of available partners (4×10^7) for the VH-B (or V κ -d) domain, and the promiscuous nature of the domain. The V κ genes were mainly ox-like and the VH genes mainly group 1, but the only examples of V κ ox1 (V κ -h, -p, -q and -r) encode Trp 91, and the VH-CDR3 motif Asp-X-Gly-X-X now predominates. Thus some features of the pHox hybridomas seem to emerge more strongly in the hierarchical library. The new partners differ from each other mainly by small alterations in the CDRs, indicating that much of the subtle diversity had not been tapped by the original random combinatorial library. More generally we find that a range of related antibodies can be made by keeping one of the partners fixed and varying the other, and this could be invaluable for fine tuning of antibody affinity and specificity.

To determine the range of antibody affinities for pHox, we recloned the combinations of VH-B with V κ -b and V κ -d (which gave weak and strong binding signals to pHox in ELISA) for secretion as soluble scFv fragments (Fig. 3, legend). Fluorescence quench titrations determined the K_d of VH-B/V κ -d for pHox-GABA as 1.0×10^{-8} M (Fig. 3a), indicating that antibodies with affinities representative of the secondary response can be selected from phage display libraries. Indeed of anti-pHox hybridomas from the secondary response, only two (out of 11 characterized) secrete antibodies of a higher affinity than VH-B/V κ -d (ref. 16). The K_d of VH-B/V κ -b for pHox-GABA was determined as 1.8×10^{-5} M (Fig. 3b); thus phage bearing

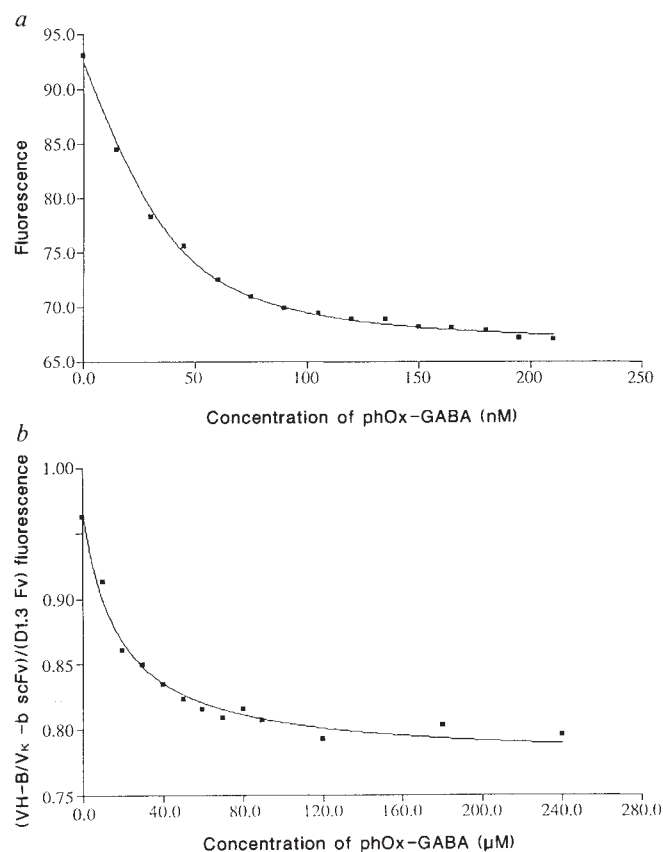
scFv fragments with weak affinities can also be selected with antigen, probably because of the avidity of the multiple antibody heads on the phage.

A second round of selection of the original, random combinatorial library from immune mice resulted in 93% of eluted clones binding pHox (Table 1a). Most of these clones were V κ -d combinations, and bound strongly to pHox in ELISA (data not shown). Few weak binders were seen. This suggested that affinity chromatography had not only enriched for binders, but also for the best. To confirm this we mixed the phage VH-B/V κ -d with a 20-fold excess of the phage VH-B/V κ -b and subjected the mixture to rounds of selection: after only two rounds, essentially all the eluted phage were VH-B/V κ -d (Table 1c).

We also constructed a random combinatorial library (2×10^6 members) from unimmunized mice, but found no pHox-binding clones after two rounds of selection (Table 1a). Immunization therefore seems to be necessary to create and/or enrich for VH or V κ domains with at least some of the features required for hapten binding. With libraries of this size ($\sim 10^6$ members), such domains need to be represented at a high frequency to reconstitute a binding site¹, and immunization ensures this by biasing the spleen lymphoid cell population heavily towards messenger RNA-rich blast cells making specific antibody (R. Hawkins and G.W., unpublished data). With larger libraries ($>10^7$) now accessible using selection³ rather than screening⁵⁻⁸, immunization may be unnecessary for the isolation of antibody fragments. It has been estimated that a library of 10^7 different antibodies will probably recognize $>99\%$ of epitopes with a dissociation constant of $\geq 10^{-5}$ M (ref. 18), and we have shown here that we can recover antibody fragments with such affinities

FIG. 3 Fluorescence quench titration of soluble scFv fragments. a, The K_d ($1.0 \pm 0.2 \times 10^{-8}$ M) for clone VH-B/V κ -d was determined by fluorescence quench titration³⁴ of purified scFv (100 nM) with 4- γ -amino-butyric acid methylene 2-phenyl-oxazol-5-one (pHox-GABA). Excitation was at 280 nm, emission was monitored at 340 nm and the K_d calculated as in refs 35 and 36. All values were calculated with standard errors included. The K_d was determined to be $1.0 \pm 0.2 \times 10^{-8}$ M with 0.38 ± 0.05 binding sites per scFv molecule. b, For measurement of the K_d of the low affinity clone VH-B/V κ -b, 2 μ M purified scFv protein was titrated with pHox-GABA as above. But to minimize light absorption by the higher concentrations of pHox-GABA required, excitation was at 260 nm and emission was monitored at 304 nm. In addition, the fluorescence values were divided by those from a parallel titration of the lysozyme binding Fv fragment D1.3 (ref. 31). The K_d was calculated as described in refs 34 and 36 and determined to be $1.8 \pm 0.3 \times 10^{-5}$ M, with a fractional quench of 0.20 ± 0.01 .

METHODS. Clones VH-B/V κ -b and VH-B/V κ -d were reamplified with VK4FOR-NotI and VH1BACK-SfiI (5'-CATGCCATGACTCGCGGCCAGCCGGCCATGGCC(G/C)AGGT(C/G)(A/C)(A/G)CTGCAG(C/G)AGTC(A/T)GG-3'), a primer that introduces an SfiI site (underlined) at the 5' end of the VH gene. VH-B/V κ -d was cloned into the phagemid pJM1 (A.D.G. and J. Marks, unpublished results) as an SfiI-NotI cassette, downstream of the pelB leader for periplasmic secretion³⁷, with a C-terminal peptide tag for detection^{31,38}, and under the control of a λ P₁ promoter³⁹. Cultures (10 l) of *Escherichia coli* N4830-1 (ref. 40) harbouring each phagemid were induced²⁶ and supernatants precipitated with 50% ammonium sulphate. The resuspended precipitate was dialysed into PBS, 0.2 mM EDTA (PBSE), loaded onto a 1.5-ml column of pHox-Sepharose⁴¹ and the column washed sequentially with 100 ml PBS; 100 ml 0.1 M Tris-HCl, 0.5 M NaCl pH 8.0; 10 ml 50 mM citrate, pH 5.0; 10 ml 50 mM citrate, pH 4.0 and 20 ml 50 mM glycine, pH 3.0. The scFv fragment was eluted with 50 mM glycine, pH 2.0, neutralized with Tris base and dialysed against PBSE. VH-B/V κ -b was cloned into a phagemid vector (A.D.G., unpublished results) based on pUC119 (ref. 42) encoding identical signal and tag sequences to pJM1, and expression induced at 30 °C in a 10-culture of *E. coli* TG1 (ref. 28) harbouring the phagemid, as in ref. 43. The low affinity of clone VH-B/V κ -b made its purification on pHox-Sepharose impossible. Therefore after concentration by ultrafiltration (Filtron, Flowgen) the supernatant (100 ml of 600 ml) was loaded onto a 1-ml column of protein A-Sepharose coupled⁴⁴ to the monoclonal antibody 9E10 that recognizes



the C-terminal peptide tag^{31,38}. The column was washed with 200 ml PBS and 50 ml PBS, 0.5 M NaCl. The scFv fragment was eluted with 100 ml 0.2 M glycine, pH 3.0, with neutralization and dialysis as before.

from phage display libraries. The antibody fragments could be rebuilt from their genes into complete antibodies, and expressed in myeloma cells if required, as described in ref. 19.

It may be possible to retain the original V_H/V_K pairings of the splenocytes, as in hybridoma technology. In principle, PCR assembly could be used to construct such 'natural' libraries, if the V genes from individual cells could be amplified and assembled in capsules. More immediately, affinity selection from combinatorial and hierarchical libraries promises an attractive route to high-affinity antibodies, in particular those from humans that are difficult to produce by hybridoma technology¹. But the use of phage display libraries is not limited to antibodies: it offers a powerful and general method to change and refine the properties of any other protein²⁰ or peptides^{21–23} that can be displayed on the phage surface. □

Received 19 March; accepted 9 July 1991.

- Winter, G. & Milstein, C. *Nature* **349**, 293–299 (1991).
- Smith, G. P. *Science* **228**, 1315–1317 (1985).
- McCafferty, J., Griffiths, A. D., Winter, G. & Chiswell, D. J. *Nature* **348**, 552–554 (1990).
- Hoogenboom, H. R. et al. *Nucleic Acids Res.* **19**, 4133–4137 (1991).
- Huse, W. D. et al. *Science* **246**, 1275–1281 (1989).
- Caton, A. J. & Koprowski, H. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6450–6454 (1990); correction, **88**, 1590 (1991).
- Mullinax, R. L. et al. *Proc. natn. Acad. Sci. U.S.A.* **87**, 8095–8099 (1990).
- Persson, M. A. A., Caotien, R. H. & Burton, D. R. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2432–2436 (1991).
- Saiki, R. K. et al. *Science* **239**, 487–491 (1988).
- Horton, R. M., Hunt, H. D., Ho, S. N., Pullen, J. K. & Pease, L. R. *Gene* **77**, 61–68 (1989).
- Huston, J. S. et al. *Proc. natn. Acad. Sci. U.S.A.* **85**, 5879–5883 (1988).
- Glockshuber, R., Malia, M., Pflitzinger, I. & Plückthun, A. *Biochemistry* **29**, 1362–1367 (1990).
- Dildrop, R. *Immun. Today* **5**, 85–86 (1984).
- Kabat, E. A., Wu, T. T., Reid-Miller, M., Perry, H. M. & Gottesman, K. S. *Sequences of Proteins of Immunological Interest* (US Department of Health and Human Services, US Government Printing Office, 1987).

- Hudson, N. W., Mudgett-Hunter, M., Panka, D. J. & Margolies, M. N. *J. Immunol.* **139**, 2715–2723 (1987).
- Berek, C., Griffiths, G. M. & Milstein, C. *Nature* **316**, 412–418 (1985).
- Alzari, P. M. et al. *EMBO J.* **9**, 3807–3814 (1990).
- Perelson, A. S. *Immunol. Rev.* **110**, 5–33 (1989).
- Orlandi, R., Güssow, D. H., Jones, P. T. & Winter, G. *Proc. natn. Acad. Sci. U.S.A.* **86**, 3833–3837 (1989).
- Bass, S., Greene, R. & Wells, J. A. *Proteins* **8**, 309–314 (1990).
- Cwirla, S. E., Peters, E. A., Barrett, R. W. & Dower, W. J. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6378–6382 (1990).
- Devlin, J. J., Panganiban, L. C. & Devlin, P. E. *Science* **249**, 404–406 (1990).
- Scott, J. K. & Smith, G. P. *Science* **249**, 386–390 (1990).
- Sambrook, J., Fritsch, E. F. & Maniatis, T. *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1989).
- Dower, W. J., Miller, J. F. & Ragsdale, C. W. *Nucleic Acids Res.* **16**, 6127–6145 (1988).
- Nagai, K. & Thøgersen, H. C. *Meth. Enzym.* **153**, 461–481 (1987).
- Mäkelä, O., Kaartinen, M., Pelkonen, J. L. T. & Karjalainen, K. *J. exp. Med.* **148**, 1644–1660 (1978).
- Gibson, T. *thesis*, Univ. Cambridge (1984).
- Gherardi, E., Pannell, R. & Milstein, C. *J. Immunol. Meth.* **126**, 61–68 (1990).
- Clackson, T., Güssow, D. & Jones, P. T. in *PCR, a Practical Approach* (eds McPherson, M. J., Taylor, G. R. & Quirke, P.) (IRL, Oxford, in the press).
- Ward, E. S., Güssow, D., Griffiths, A. D., Jones, P. T. & Winter, G. *Nature* **341**, 544–546 (1989).
- Güssow, D. & Clackson, T. *Nucleic Acids Res.* **17**, 4000 (1989).
- Sanger, F., Nicklen, S. & Coulson, H. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
- Eisen, H. N. *Meth. med. Res.* **10**, 115–121 (1964).
- Segal, I. *Enzyme Kinetics* 73–74 (Wiley-Interscience, New York, 1975).
- Foot, J. & Milstein, C. *Nature* (in the press).
- Better, M., Chang, C. P., Robinson, R. R. & Horwitz, A. H. *Science* **240**, 1041–1043 (1988).
- Munro, S. & Pelham, H. R. B. *Cell* **46**, 291–300 (1986).
- Shimatake, H. & Rosenberg, M. *Nature* **292**, 128–132 (1981).
- Gottesman, M. E., Adhya, S. & Das, A. J. *molec. Biol.* **140**, 57–75 (1980).
- McManus, S. & Riechmann, L. *Biochemistry* **30**, 5851–5857 (1991).
- Vieira, J. & Messing, J. *Meth. Enzym.* **153**, 3–11 (1987).
- De Bellis, D. & Schwartz, I. *Nucleic Acids Res.* **18**, 1311 (1990).
- Harlow, E. & Lane, D. *Antibodies: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1988).

ACKNOWLEDGEMENTS. We thank J. McCafferty for help in constructing the vector fdDOG1; J. Marks for the vector pJM1; C. Marks for oligonucleotide primers for cDNA synthesis and for D1.3 Fv protein; E. Gherardi for spleen mRNA; J. Foot for help with fluorescence quench titrations; and M. Dreher, L. Riechmann, J. Jarvis, D. Chiswell, A. R. Fersht and C. Milstein for advice. Supported by the MRC, the D. Collier Research Foundation, Leuven, and the European Molecular Biology Organisation (H.R.H.), and by the Louis Jeantet Foundation (A.D.G. and T.C.).

Phosphorylation-regulated Cl[−] channel in CHO cells stably expressing the cystic fibrosis gene

Joseph A. Tabcharani*, Xiu-Bao Chang†, John R. Riordan† & John W. Hanrahan*‡

* Department of Physiology, McGill University, 3655 Drummond Street, Montréal, Québec H3G 1Y6, Canada

† Research Institute, Hospital for Sick Children, and University of Toronto, Toronto, Ontario M5G 1X8, Canada

A CYCLIC AMP-stimulated chloride conductance appears when the cystic fibrosis gene is expressed in non-epithelial cells by infection with recombinant viruses^{1,2}. Cyclic AMP-stimulated conductance in this system is mediated by the same ohmic, low-conductance Cl[−] channel as in human secretory epithelia^{2–4}, but control of this channel by phosphorylation has not been directly demonstrated. Here we report the appearance of the low-conductance Cl[−] channel in Chinese hamster ovary cells after stable transfection with the cystic fibrosis gene. The channel is regulated on-cell by membrane-permeant analogues of cAMP and off-cell by protein kinases A and C and by alkaline phosphatase. These results are further evidence that the cystic fibrosis transmembrane regulator is a Cl[−] channel which can be activated by specific phosphorylation events and inactivated by dephosphorylation; they reveal an unsuspected synergism between converging kinase regulatory pathways.

The coding sequence of the cystic fibrosis transmembrane regulator (CFTR) was cloned behind the metallothionein pro-

motor of a plasmid that also contained a mutant dihydrofolate reductase gene, driven by the simian virus 40 early promoter (Fig. 1a). Stably transformed colonies were selected with methotrexate after calcium phosphate transfection of Chinese hamster ovary (CHO)-K1 cells. CFTR-expressing variants were chosen for further study on the basis of their capacity to produce a protein of the same apparent size as that present in T84 cell membranes in western blots probed with monoclonal antibodies against CFTR (Fig. 1b). CFTR protein was localized in a highly enriched plasma membrane vesicle fraction. In variants containing nearly the same amount of CFTR as T84 cells, cAMP-regulated chloride permeability, as monitored by ¹²⁵I efflux, was indistinguishable from that in T84 cells (Fig. 1c).

Patch-clamp recording was used to identify the channel responsible for cAMP-stimulated ¹²⁵I efflux. Channels became active in cell-attached patches after a lag of 69 ± 26 seconds when cells were exposed to membrane-permeant derivatives of cAMP; this was reversed by washing cAMP from the bath (*n* = 7, Fig. 2a). The channel was observed in 80% of all seals during cAMP stimulation (225/282) at an average density of between five and ten channels per patch. By contrast, it was recorded only once in 55 patches on unstimulated, CFTR-transfected cells and was never observed on cAMP-stimulated CHO cells that had been transfected with vector alone (0/31). Figure 2b shows that open probability was relatively independent of voltage, despite increased flickering at hyperpolarized potentials. Flickering was not observed using excised patches (see below), therefore these brief closures may reflect voltage-dependent, fast channel block by some anion in the cytosol. The current-voltage relationship rectified slightly in the outward direction during cell-attached recordings (Fig. 2c), but was linear (*r*² = 0.9997) when patches were excised and bathed symmetrically with 154 mM Cl[−] (data not shown). In cell-attached patches the reversal potential (*E*_{rev}) was near the membrane potential (0.6 ± 0.3 mV applied potential) and the slope conductance at *E*_{rev} was 9.6 ± 0.5 pS (*n* = 5). The *E*_{rev} shifted to +32.4 ± 2.0 mV when the pipette solution contained 110 mM sodium gluconate and

‡ To whom correspondence should be addressed.

40 mM NaCl, indicating a $P_{Cl}:P_{cation}$ ratio of ~ 14 . Activity disappeared within 2 min after patches were excised (102/102; upper trace, Fig. 2d). This rundown occurred with okadaic acid (10 μ M), a potent inhibitor of type 1 and 2A protein phosphatases⁵, and Mg-ATP in the bath, but not with cAMP-dependent protein kinase catalytic subunit (45.1–180.5 nM) and Mg-ATP (lower trace, Fig. 2d). Figure 2e shows the effect of bath protein kinase A (PKA) on the number of patches having active channels after excision as a function of time. The decline in activity without PKA was well fitted by two exponentials having time constants of 0.4 and 8.6 s, respectively. The level of activity apparently depends on a balance between phosphorylation and dephosphorylation so that excision in the absence of PKA shifts the channel towards dephosphorylation and deactivation.

Channels that became quiescent could be reactivated by exposing the cytoplasmic side to PKA in the presence of Mg-ATP (Fig. 3a). PKA increased the open probability (P_o) from zero to 0.407 ± 0.023 ($n = 9$). Exposure to protein kinase C (PKC) increased P_o in 15/16 patches, but this effect, which was quantified using 11 patches, was small compared with the stimulation elicited by PKA (P_o increasing from 0 to 0.034 ± 0.010 ; Fig. 3b). Interestingly, both the rate and magnitude of stimulation by PKA were increased in the presence of PKC compared with PKA when it was added alone (Fig. 3c). The mechanism underlying this potentiation by PKC is unknown, but CFTR has a highly charged regulatory domain (the R domain) bearing nine consensus sequences for phosphorylation by PKA and seven

for PKC⁶: phosphorylation by PKC may expose other PKA sites, perhaps by altering the conformation of the R domain.

Open probability increased slowly from 0.3 to ~ 0.6 after PKA activation when the bath also contained fluoride (10 mM) or metavanadate (10 mM), which are widely used nonspecific inhibitors of protein phosphatases⁷. This suggests some endogenous phosphatase activity may remain associated with the membrane when patches are excised, as suggested in an earlier study of K^+ channels in *Aplysia* neurons⁸. It is not yet clear which phosphatase deactivates the channel in our patches: P_o increased normally in the presence of okadaic acid (1–10 μ M), which excludes type 1 and type 2A protein phosphatases⁵ and in the nominal absence of calcium or calmodulin, which excludes type 2B (refs 7, 9). The spontaneous increase in P_o after excision was sensitive to fluoride, which argues that type 2C protein phosphatase⁷ is also not responsible.

To determine whether channels could be deactivated in excised patches by an exogenous phosphatase, we activated the Cl^- channel using PKA and then added intestinal alkaline phosphatase (EC 3.1.3.1). Alkaline phosphatase (95 U per ml) decreased P_o by more than 90% within 2 min in the continued presence of 180.5 nM PKA (Fig. 4). Deactivation of the channel by exogenous alkaline phosphatase was insensitive to okadaic acid (10 μ M), completely blocked by fluoride (10 mM) or metavanadate (10 mM), and partially inhibited by orthovanadate (100 μ M). Thus exogenous alkaline phosphatase had the same inhibitor sensitivities as the endogenous phosphatase, which

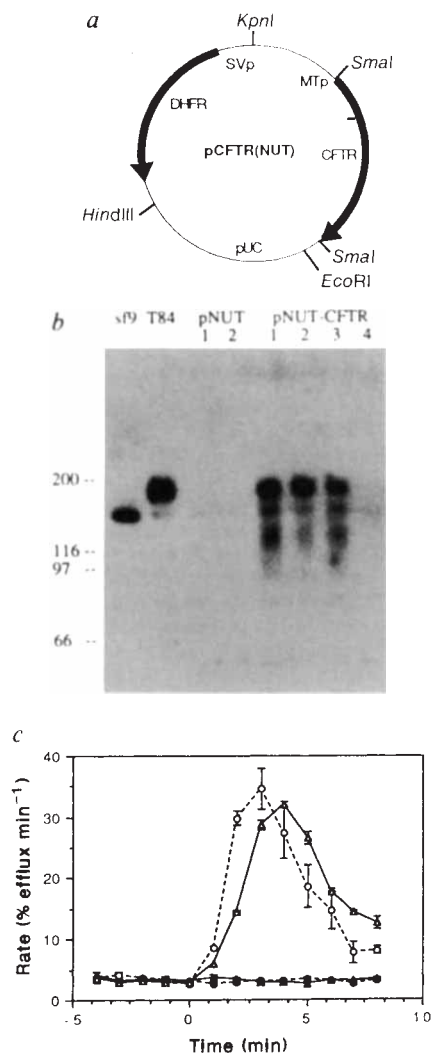


FIG. 1 Functional expression of CFTR in CHO cells stably transfected with the cystic fibrosis gene. *a*, The recombinant plasmid was constructed starting from a plasmid, pNUT, prepared by Palmiter *et al.*¹⁸. Removal of the human growth hormone gene that it originally contained left a *SmaI* site, immediately downstream of a mouse metallothionein promoter (MT_p), and human growth hormone polyadenylation sequence. The CFTR coding sequence (including nucleotides 72 to 4,721 of the complete cDNA⁶) was inserted into the *SmaI* site in the sense orientation. As indicated, the original vector (pNUT) also contains a mutant dihydrofolate reductase (DHFR) gene under the control of the SV40 early promoter (SV_p). *b*, The presence of immunoreactive CFTR protein in 3 of 4 variants selected independently in methotrexate (10 μ M) was detected by western blotting of crude membrane preparations from these cells. Migration of M_r ($\times 10^{-3}$) markers is shown on the left. The principal band is of the same apparent molecular weight as the native protein endogenously expressed in T84, and larger than the minimally glycosylated form produced by insect Sf9 cells infected with a CFTR-containing baculovirus². No bands were detected in cells in which plasmid vector alone was integrated (pNUT 1 and 2). *c*, The effect of cAMP (added at time 0) on the rate of ^{125}I efflux in CFTR-expressing CHO variant 1 (open circles) and in T84 cells (open triangles). The filled symbols represent data from corresponding cells not treated with cAMP.

METHODS. CHO-K1 cells were cultured at 37 °C in 5% CO₂. Subconfluent cells were transfected with pNUT-CFTR in the presence of 20 mM HEPES, pH 7.05, 137 mM NaCl, 5 mM KCl, 0.7 mM Na₂HPO₄, 6 mM dextrose and 125 mM CaCl₂. About 24 h later, 10 μ M methotrexate was added to the medium. Cells continued to grow in the selective medium (10 μ M methotrexate) for about 10 days. Individual colonies were picked and amplified and crude membrane was harvested from the amplified cells. The crude membrane proteins were subjected to SDS-PAGE, and a western blot of the gel was probed with a CFTR monoclonal antibody. ^{125}I was loaded in to the cells in efflux buffer (136 mM NaCl, 3 mM KCl, 2 mM CaCl₂, 11 mM glucose and 20 mM HEPES, pH 7.4). Extracellular ^{125}I was washed away and cells were equilibrated for 60 s in a final 0.5-ml aliquot. The first four aliquots were used to establish a stable baseline in efflux buffer alone. Agonists (10 μ M forskolin, 200 μ M dibutyryl cAMP, and 1 mM IBMX, dissolved in dimethylsulphoxide) were added at time zero. The same amount of dimethylsulphoxide without agonists was added to the efflux buffer in the control samples.

apparently remains associated with the patch. Indeed, the endogenous phosphatase in these CHO patches may be an alkaline phosphatase; however, more specific protein phosphatases could be involved in the epithelial cells to which the channel is native. Our results contrast with those in excitable cells: three protein phosphatases (types 1, 2A and 2B) are implicated in the regulation of the L-type calcium channel in heart cells¹⁰, and protein phosphatase types 1 and 2A both

regulate distinct K^+ channels from snail neurons incorporated into lipid bilayers¹¹.

Single Cl^- channels that appear in CFTR-expressing CHO cells are indistinguishable from an ohmic, low-conductance channel in human pancreatic duct³ and T84 (ref. 4) cells and baculovirus-infected insect cells overexpressing CFTR². The observation that CFTR expression generates this same channel in different systems that do not normally have cAMP-regulated

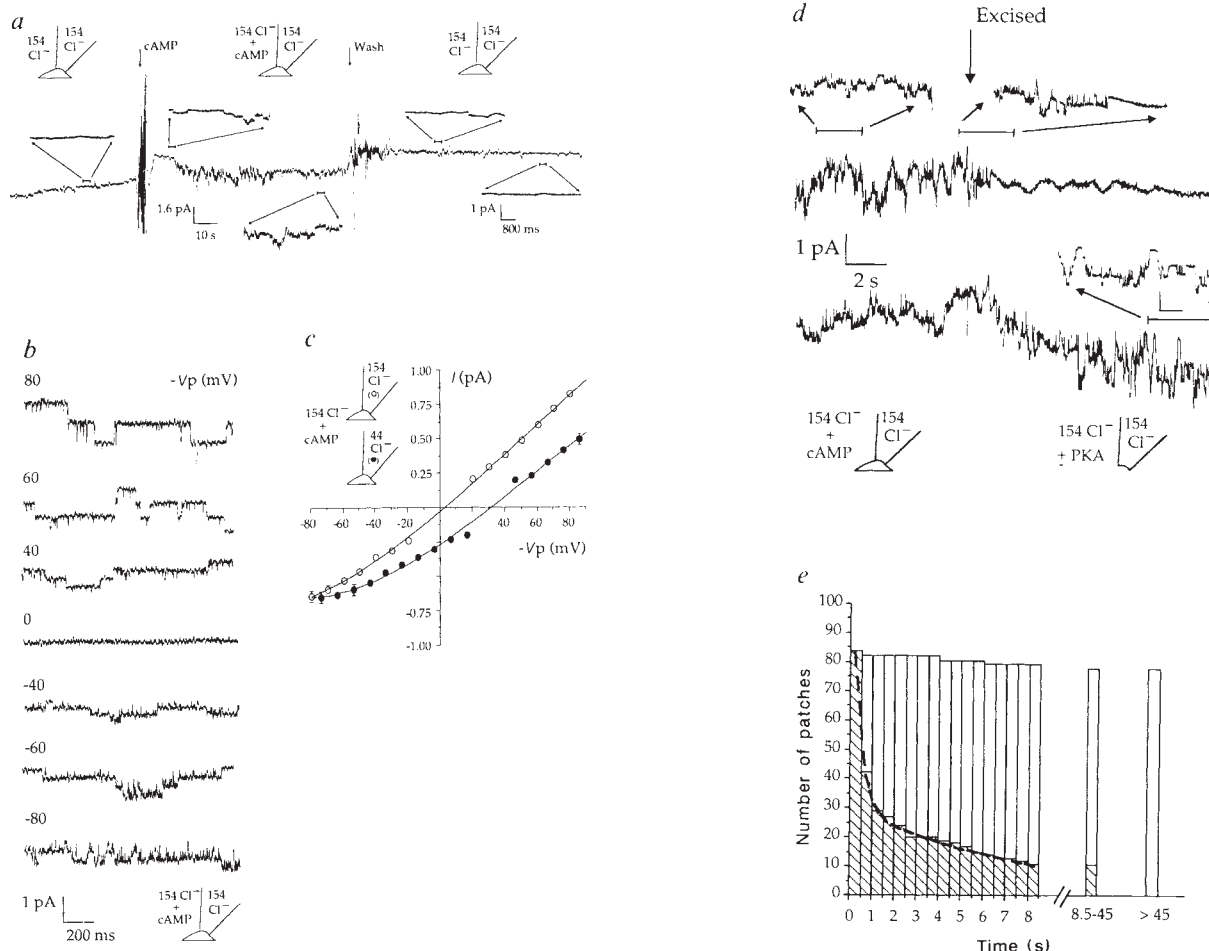


FIG. 2 Low-conductance chloride channels activated by cAMP in cell-attached patches and their deactivation by patch excision. **a**, Continuous recording on a CFTR-transfected CHO cell showing channel activity before cAMP addition, during stimulation, and after cAMP was washed from the bath. The patch potential was clamped at -40 mV (hyperpolarized) with respect to the resting membrane voltage in this and all subsequent cell-attached recordings. Insets show segments at expanded scale. Activation was not caused by inadvertently excising the patches because, as shown below, excision under these conditions causes the channel to deactivate. This channel was not observed in any patches from control nontransfected CHO cells. **b**, Representative traces at different holding potentials in the cell-attached configuration. **c**, Mean current-voltage relationships determined in cell-attached patches with pipette solutions containing 150 mM NaCl (○), or 40 mM NaCl + 110 mM sodium gluconate (●). **d**, Effect of excising patch into solution with (upper trace) and without (lower trace) cAMP-dependent protein kinase catalytic subunit (PKA) in the bath. The bath also contained 0.5 mM ATP and 2 mM Mg^{2+} in both cases. Scale for insets: 1 pA vertical, 400 ms horizontal. **e**, Number of patches with active channels as a function of time after excision in the absence (hatched bars) or presence (open bars) of PKA in the bath. Dashed line indicates least-squares best fit by two exponentials having time constants of 0.37 and 8.58 s and relative areas of 0.25 and 0.75, respectively.

METHODS. Cells were plated at low density on glass coverslips and cultured at 37 °C in 5% CO_2 for 1-4 days before use. All experiments were at 37 °C. The pipette solution contained (in mM) 150 NaCl, 2 $CaCl_2$, 10 TES (*N*-

tris(hydroxymethyl)methyl-2-aminoethane sulphonic acid) (pH 7.4). The bath solution contained (in mM) 140 NaCl, 4 KCl, 2 $MgCl_2$, 1 EGTA, 0.5 $CaCl_2$, 5 glucose, 10 TES (pH 7.4), except when testing the effect of PKA on deactivation, when 0.5 mM ATP and purified catalytic subunit of cAMP-dependent protein kinase (PKA; 45.1-180.5 nM) were added to the bath. Type II bovine cardiac PKA (prepared in the laboratory of M. P. Walsh), had a specific activity of 0.45 μ mol phosphate incorporated per min per mg enzyme when assayed at 30 °C in 20 mM Tris-HCl, pH 7.5, containing 1 mM EGTA, 5 mM $MgCl_2$ and 0.2 mM ATP, using histone IIA (Sigma) as substrate. Low- Cl^- pipette solution was prepared by replacing 110 mM NaCl with 110 mM sodium gluconate. The cAMP mixture contained (final concentration in μ M): 500 dibutyl cAMP, 10 forskolin and 10 IBMX. Identical results were obtained when 500 μ M 8-(4-chlorophenylthio)-cAMP was used instead of mixture. The chamber (500 μ l) was perfused with >6 ml solution to wash out cAMP. Standard patch clamp techniques were used in this and subsequent figures. The mean number of channels open ($\langle I \rangle / i = NP_o$) was computed for each 10-s segment of the record. Open probability P_o was estimated by assuming the number of channels in the patch N to be the largest number of simultaneous openings observed in long records (>2 min) during stimulation. The maximum number of steps observed (and hence estimated channel number) was similar on- and off-cell, and estimates of N were never higher after excision even though P_o did increase under some conditions. Recordings were low-pass filtered at 150 Hz and digitized at 1 kHz. Other details of data acquisition and analysis have been described⁴.

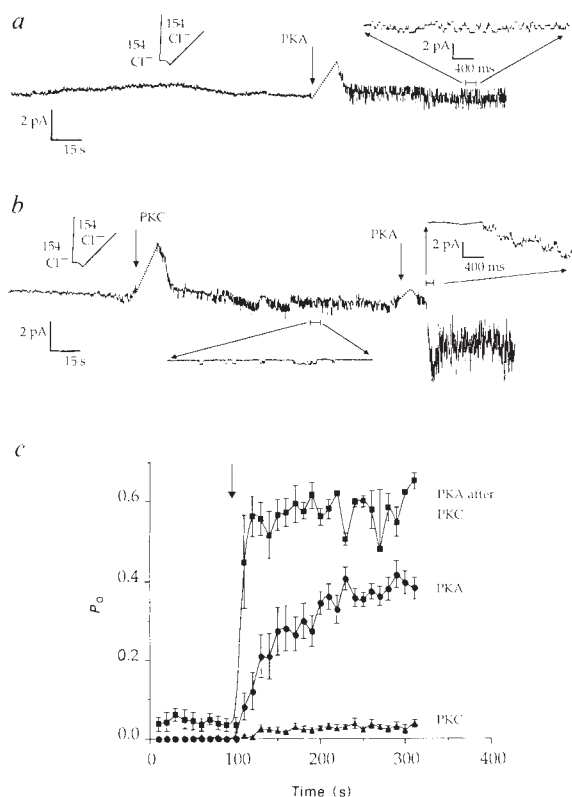


FIG. 3 Effect of adding protein kinases on Cl^- channel open probability in patches isolated from CHO cells that express CFTR. *a*, Activation of Cl^- channels by simultaneous addition of 180.5 nM PKA and 0.5 mM ATP in the presence of 2 mM MgCl_2 . This stimulation was abolished by PKA $^-$ inhibitor peptide (6–22)amide (4–40 μM ; ref. 19). Adding 0.5 mM ATP alone under these conditions had no effect. *b*, Activation by 3.78 nM PKC, 5 μM 1,2-dioctanoylglycerol (8:0) and 0.5 mM ATP in the presence of 2 mM MgCl_2 . This patch contained at least 12 channels. Note potentiated PKA response after PKC preincubation. The weaker potency of PKC was not due to its lower concentration because increasing the amount of enzyme 8-fold caused no additional stimulation. Potentiation was specific and did not occur at the level of PKA, for example through phosphorylation by PKC of the cAMP-dependent protein kinase catalytic subunit, because it was not observed when PKA, PKC, DiC_8 , and Mg-ATP were preincubated together under similar conditions and then added to the bath containing 20 μM protein kinase C-inhibitor peptide (19–36; ref. 20). *c*, Effect of adding protein kinases A and C on open-state probability. Effect of PKA ($n=9$), PKC ($n=11$), and of PKA after exposure to PKC ($n=7$); means $\pm$ s.e.

METHODS. Rat brain PKC II, as identified by molecular weight and western blotting with a monoclonal antibody, was obtained from the laboratory of M. P. Walsh. It had a specific activity of 0.25 μmol phosphate incorporated $\text{min}^{-1} \text{mg}^{-1}$ enzyme at 30 $^\circ\text{C}$ when assayed in 20 mM Tris-HCl buffer, pH 7.5, containing 10 mM MgCl_2 , 0.1 mM CaCl_2 , 0.3 mg ml^{-1} L- α -phosphatidyl-L-serine, 60 $\mu\text{l ml}^{-1}$ 1,2-diolein, 0.03% Triton X-100 and 10 μM ATP and using calf thymus lysine-rich histone IIIA (Sigma) as substrate.

Cl^- permeability further strengthens the hypothesis that this channel is the cystic fibrosis gene product. *In vitro* phosphorylation of CFTR by PKA is well established¹², but we have directly demonstrated that the low-conductance Cl^- channel is regulated

by kinases and phosphatases. The potentiating effect of PKC on PKA activation was unexpected, although PKC agonists stimulate Cl^- secretion across some epithelia^{13,14}. There are precedents for dual regulation of channels by kinases: the dihydropyridine receptor is a substrate for PKA and PKC, both of which stimulate L-type Ca^{2+} currents in cardiac cells^{15,16}, although the kinases preferentially phosphorylate different subunits of the dihydropyridine receptor¹⁷ and potentiation has not been reported. Studying the molecular details of kinase regulation by genetic manipulations and patch clamping should help to elucidate the mechanism by which the R domain controls CFTR function, and the general problem of how converging kinase regulatory paths interact at a target protein. □

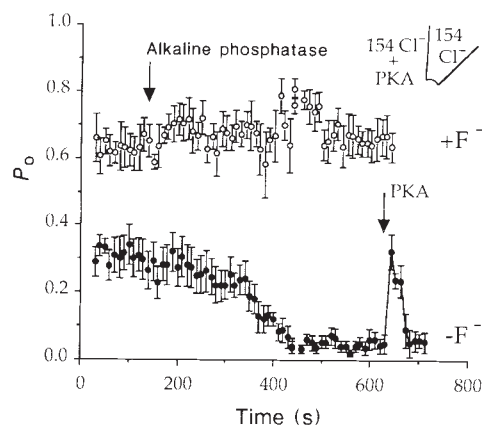


FIG. 4 Deactivation of Cl^- channels in excised patches following exposure to alkaline phosphatase. Addition of alkaline phosphatase from bovine intestinal mucosa (95 U ml^{-1} , where 1 unit hydrolyses 1 μmol *p*-nitrophenyl phosphate per min at pH 9.8 and 37 $^\circ\text{C}$; Sigma) reduced P_o of channels that had been activated by PKA from 0.3 to 0.03 (●; mean $\pm$ s.e., $n=7$). Incomplete deactivation was probably not due to the relatively low pH (7.4) or to the low concentrations of enzyme and zinc used because similar results were obtained at pH 8.0, when enzyme concentration was doubled, and when 100 μM Zn^{2+} was added to bath. Increasing the concentration of PKA to 361 nM in the continued presence of alkaline phosphatase briefly restored P_o to the pre-phosphatase level, probably owing to inadequate mixing ($n=2$). Inhibition by alkaline phosphatase was blocked when the bath contained 10 mM NaF (open circles; mean $\pm$ s.e., $n=9$).

Received 6 June; accepted 2 July 1991.

- Anderson, M. P., Rich, D. P., Gregory, R. J., Smith, A. E. & Welsh, M. J. *Science* **251**, 679–682 (1991).
- Kartner, N. et al. *Cell* **64**, 681–691 (1991).
- Gray, M. A., Harris, A., Coleman, L., Greenwell, J. R. & Argent, B. E. *Am. J. Physiol.* **257**, C240–C251 (1989).
- Tabcharani, J. A., Low, D., Elie, D. & Hanrahan, J. W. *FEBS Lett.* **270**, 157–164 (1990).
- Bialojan, C. & Takai, A. *Biochem. J.* **256**, 283–290 (1988).
- Riordan, J. R. et al. *Science* **245**, 1066–1073 (1989).
- Shenolikar, S. & Nairn, A. C. *Advances in Cyclic Nucleotide Protein Phosphorylation Research* **23**, 1–121 (1991).
- Shuster, M. J., Camardo, J. S., Siegelbaum, S. A. & Kandel, E. R. *Nature* **313**, 392–395 (1985).
- Cohen, P. A. *Rev. Biochem.* **58**, 453–508 (1989).
- Herschler, J., Kameyama, M., Trautwein, W., Mieskes, G. & Soeling, H. D. *Eur. J. Biochem.* **165**, 261–266 (1987).
- Levitani, I. B., Chung, S. & Reinhardt, P. H. *Advances in Second Messenger and Phosphoprotein Research* **24**, 36–40 (1990).
- Gregory, R. J. et al. *Nature* **347**, 382–386 (1990).
- Chang, E. B., Wang, N. S. & Rao, M. C. *Am. J. Physiol.* **249**, C356–C361 (1985).
- Fondacaro, J. D. & Henderson, L. S. *Am. J. Physiol.* **249**, G422–G426 (1985).
- Brum, G., Flockerzi, V., Hofmann, F., Osterreider, W. & Trautwein, W. *Pfluegers Arch. ges. Physiol.* **398**, 147–154 (1983).
- Lacerda, A. E., Rampe, D. & Brown, A. M. *Nature* **335**, 245–251 (1988).
- Jahr, H., Nastainczyk, W., Röhrkasten, A., Schneider, T. & Hofmann, F. *Eur. J. Biochem.* **178**, 535–542 (1988).
- Palmiter, R. D. et al. *Cell* **50**, 435–443 (1987).
- Glass, D. B. et al. *J. Biol. Chem.* **264**, 8802–8810 (1989).
- Hardie, G. *Nature* **335**, 592–593 (1988).

ACKNOWLEDGEMENTS. This work was supported by the US and Canadian Cystic Fibrosis Foundations, the NIH and MRC (Canada), and the Respiratory Health Network of Centres of Excellence. J.W.H. is an MRC Research Scholar.

New antiviral strategy using capsid–nuclease fusion proteins

Georges Natsoulis & Jef D. Boeke

Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, 725 North Wolfe Street, Baltimore, Maryland 21205, USA

OVEREXPRESSION of dominant-negative mutants of various viral proteins can result in 'intracellular immunization' (refs 1, 2). Here we describe a new approach to interfering with viral replication in which a nuclease is fused to a capsid component so that the nuclease is encapsidated inside the virion where it can inactivate viral nucleic acid. We used Ty1, a yeast retrotransposon whose transposition closely parallels retroviral replication mechanisms and serves as an easily manipulated model for the retroviral infection process³. We constructed fusion genes consisting of the region encoding the N-terminal portion of the TYA/TYB open reading frames of retrotransposon Ty1 and either of two different nuclease genes. Ty1–nuclease fusion proteins are targeted to Ty1 virus-like particles, and are active in degrading nucleic acids. A Ty1–barnase fusion protein causes 98–99% reduction in the efficiency of Ty1 transposition *in vivo*, presumably by degrading encapsidated Ty1 RNA. This strategy, referred to as capsid-

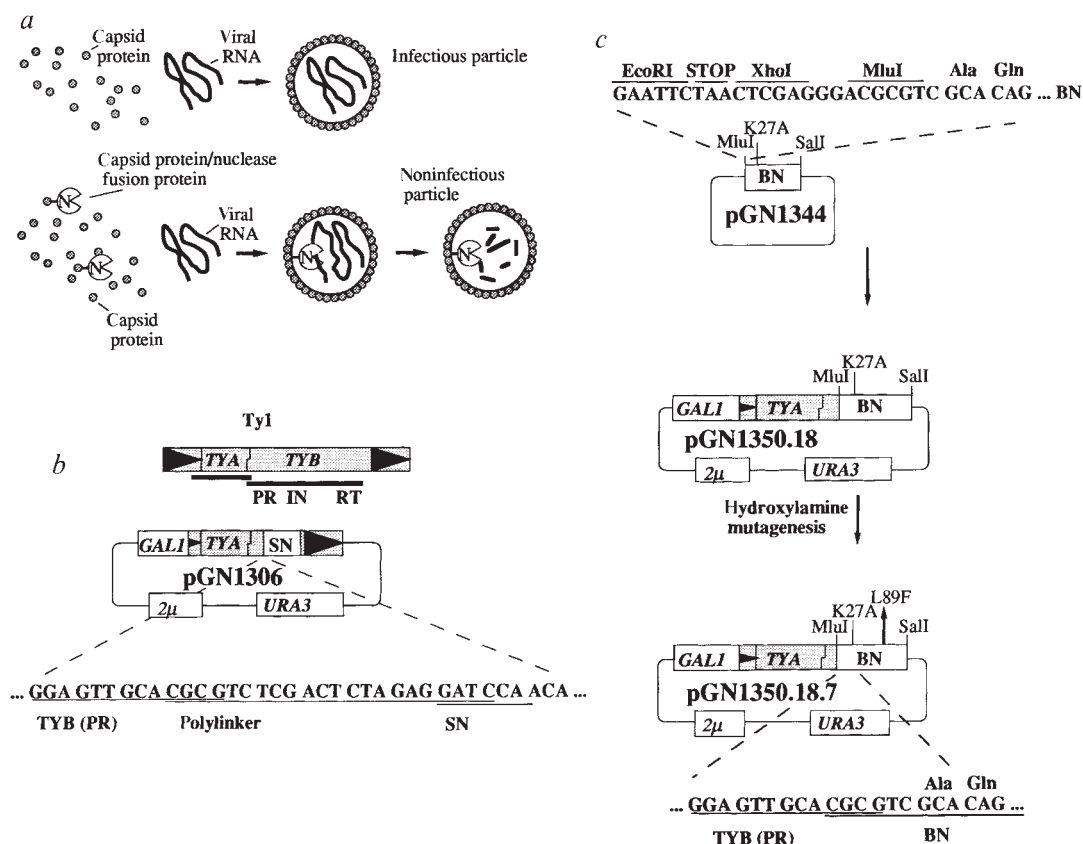
targeted viral inactivation, may be useful for interfering with the replication of retroviruses and other viruses.

Retroviruses, and the related retrotransposons, are especially good candidates for capsid-targeted viral inactivation (CTVI; Fig. 1a) because of their assembly mechanism. Retroviral capsid assembly models suggest a method for targeting nuclease molecules to the inner surface of the capsid, providing access to viral nucleic acid. Retrovirus and retrotransposon reverse transcriptases (RTs) are expressed as Gag–Pol fusions, in which Pol forms the C terminus. Mechanisms of readthrough to form Gag–Pol protein vary, including frameshifting⁴, nonsense suppression⁵ and splicing^{6,7}. In every case the RT domain, which must gain access to viral RNA, is C-terminal to Gag protein sequences. By substituting a nuclease coding region for RT, the nuclease should be assembled into the particle so as to access viral RNA. Fusion proteins encoded by chimaeric genes consisting of heterologous ORFs fused to *gag* or *pol* are directed to retroviral particles^{8,9} and Ty1 virus-like particles (VLPs)¹⁰.

One difficulty for CTVI is that nonspecific nucleases are likely to be toxic to host cells through indiscriminate nucleic-acid degradation^{11,12}. Staphylococcal nuclease¹³ (SN) and barnase (BN) are ideal candidates for antiviral activity. SN activity is completely dependent on Ca^{2+} ; BN activity can be directly suppressed by expressing the barstar peptide¹¹. Both nucleases are small, consisting of one polypeptide chain; collections of mutant nucleases that range down to 10^{-5} the specific activity of wild-type nucleases exist^{14,15}.

FIG. 1 Capsid-targeted viral inactivation strategy and plasmid constructions. **a**, Strategy for CTVI. Circles, capsid proteins; heavy line, viral nucleic acid; N, nuclease. Upper portion, normal assembly process; lower portion, CTVI. **b**, Ty1–SN fusion plasmid pGN1306. PR, protease; IN, integrase; RT, reverse transcriptase; SN, staphylococcal nuclease; 2 μ , origin of replication of yeast 2 micrometer plasmid; *GAL1*, yeast *GAL1* promoter; boxed triangles, LTRs. **c**, Ty1–BN fusion plasmids.

METHODS. **b**, The SN gene was fused to Ty1 at position 2,600 of Ty1 (ref. 27). This segment contains intact TYA as well as part of TYB (PR region). The SN gene is fused to Ty1 so that the first amino acid of SN encoded is the proline at position –5. A short sequence, derived from a polylinker, connects TYB–PR to the SN coding region. The Ty1 promoter was replaced by the yeast *GAL1* promoter. **c**, The polymerase chain reaction (PCR) was used to generate cassettes containing the BN structural gene. The PCR oligonucleotide 5' to the BN ORF consisted of an *EcoRI* site, followed by a stop codon (intended to eliminate any possible expression of the BN gene) in frame with the downstream BN gene, followed by *XhoI* and *MluI* sites for fusion of the BN gene to Ty1, followed by the first Ala codon of mature BN protein. Wild-type BN cloned in plasmid pMT416 (ref. 28) was initially used as the template. Surprisingly, when this PCR product was cloned into M13mp18 we found deleterious mutations in the BN coding region of most of the recovered recombinant phage (data not shown), suggesting that the upstream termination codon and the lack of a suitable initiation codon failed to prevent lethality of an intact BN ORF to M13-infected *E. coli*. The PCR procedure was repeated using a plasmid containing a mutation in the BN ORF as the template (Lys 27 to Ala, designated K27A). The



specific activity of BN-K27A is 1.3% of that of wild-type BN (ref. 15). The PCR product was cloned into M13mp18, resulting in pGN1344. The BN-K27A gene was fused to Ty1 at position 2,600. Although pGN1350.18 was tolerated by *E. coli*, healthy yeast transformants carrying this plasmid could not be obtained, even when the transformation mixture was plated on glucose-containing medium, suggesting that even minuscule expression of the fusion protein is highly deleterious to yeast cells. Plasmid pGN1350.18 was mutagenized with hydroxylamine, giving rise to pGN1350.18.7, which contains the additional mutation L89F, and other derivatives (Table 1).

TABLE 1 Effects of plasmids encoding Ty1-nuclease fusion proteins on Ty1 transposition

Strain number	Test plasmid (plasmid type)	Transposition fraction	(N)	Inhibition (%)	Strain number	Test plasmid (plasmid type)	Transposition fraction	(N)	Inhibition (%)
SN plasmid					BN plasmids				
GN175	pRS316* (control)	0.73	81	NA	Hydroxylamine mutagenesis experiment				
GN172	pGN1304† (SN)	0.63	90	14	GN328-330	pGN1358‡ (control)	0.45	179	NA
SN plasmid + Ca²⁺					GN334	pGN1350.18.1§ (BN)	0.045	89	90
GN175 (6 mM)	pRS316* (control)	0.64	81	NA	GN335	pGN1350.18.7 (BN)	0.0060	498	99
GN175 (100 mM)	pRS316* (control)	0.64	82	NA	GN336	pGN1350.18.9 (BN)	0.57	61	0
GN172 (6 mM)	pGN1304† (SN)	0.41	60	35	GN337	pGN1350.18.10 (BN)	0.46	91	0
GN172 (100 mM)	pGN1304† (SN)	0.36	36	43		pGN1350.18.11 (BN)	0.49	69	0
GN173 (6 mM)	pGN1306† (SN)	0.34	90	46		pGN1350.18.13 (BN)	0.46	78	0
GN173 (100 mM)	pGN1306† (SN)	0.34	97	46	Plasmid segregation experiment				
					GN349-351	none (Ura-deriv. of 328-330)	0.51	156	NA
					GN338-340, GN344	none (Ura-deriv. of 335)	0.53	123	NA
					Recloning experiment				
					GN328-330	pGN1358‡ (control)	0.35	232	NA
					GN378-380	pGN1398¶ (BN)	0.0060	287	98

All strains are diploid strains bearing plasmid pJEF1678 in addition to a test plasmid. These diploids were analysed as in Fig. 2. The diploids were constructed from single transformants each carrying one of the two plasmids; the two untransformed parent strains were JB740 (*MAT α his3 Δ 200 leu2 Δ 1 ura3-167*) and JB503 (*MAT α his3 Δ 200 trp1 Δ 1 ura3-52 lys2*). Diploids derived from at least two independent transformants were tested for each strain. Transposition fraction is the number of plasmid-free (*His*⁻) cells that become resistant to G418 during the period in which transposition is induced; G418 resistance can be caused by the insertion of one or more Ty1-*neo* elements. *N*, number of plasmid-free segregants analysed. Inhibition is defined as the reduction in transposition attributed to the inhibitory plasmid, and calculated by dividing the transposition fraction in the presence of the (potentially) inhibitory plasmid by the transposition fraction in the presence of the control plasmid. This fraction is multiplied by 100 and subtracted from 100 to give % inhibition. When the experimental transposition fraction exceeded that of the control, % inhibition was scored as 0. NA, not applicable.

* Ref. 26.

† pGN1304: *GAL1*-Ty1-SN fusion inserted in pGN621 (a *URA3*, 2 μ m vector constructed by inserting the blunt-ended 2.2-kilobase *EcoRI* origin-containing fragment of yeast 2- μ m circle DNA, B form, at the *SmaI* site of Yip5). Plasmid pGN1306: *GAL1*-Ty1-SN fusion (Fig. 1b). This construct is identical to pGN1304 except that the polylinker sequence between Ty1 and SN is slightly shorter.

‡ pGN1358: *GAL1*-Ty1 fusion truncated at the *SalI* site +2174 (ref. 27) inserted in pGN621 between the *EcoRI* and *SalI* sites.

§ This plasmid could not be recovered in *E. coli* and was not studied further.

|| Strains not saved.

¶ Plasmid pGN1398: BN-K27A, L89F insert from pGN1350.18.7 recombined into identical unmutagenized plasmid backbone.

We have fused the SN and BN genes in frame to the *TYB* (*pol*) gene of a *GAL1*-Ty1 plasmid (Fig. 1b, c). Ty1 is an ideal test system because: (1) Ty1 retrotransposition is mechanistically similar to retroviral replication¹⁶⁻²¹; (2) Ty1 VLPs are probably direct retrotransposition intermediates^{22,23}; (3) sensitive transposition assays exist^{16,24}; and (4) the yeast system allows large-scale mutant screening. The Ty1-SN fusion gene plasmid, pGN1306, contains full-length *TYA* (*gag*) as well as the intact protease domain of *TYB*. We also constructed pJEF1678, a *HIS3 GAL1*-Ty1 plasmid marked with *neo* in the Ty1 element (Fig. 2a). Plasmid pGN1306 and control plasmids were separately introduced into yeast strains bearing plasmid pJEF1678. VLPs were isolated from these strains and assayed for RT activity, nuclease activity and proteins immunoreactive to anti-TYA and anti-SN antibodies (Fig. 3). Ca²⁺-dependent nuclease activity is observed only in the pGN1306-derived VLPs. The nuclease peak cofractionates with RT activity, VLP proteins and proteins reacting with anti-SN antibody. The active SN fusion protein is efficiently targeted to VLPs.

The BN gene required modifications, due to its toxicity to *Escherichia coli* and yeast, before it could be fused to *TYB* in pGN1350.18 (Fig. 1c). This plasmid, bearing the K27A mutation¹⁵ in BN, was tolerated in *E. coli*, but extremely difficult to introduce into yeast, even on glucose medium. Thus the TYA-PR-BN-K27A fusion protein produced is very deleterious to

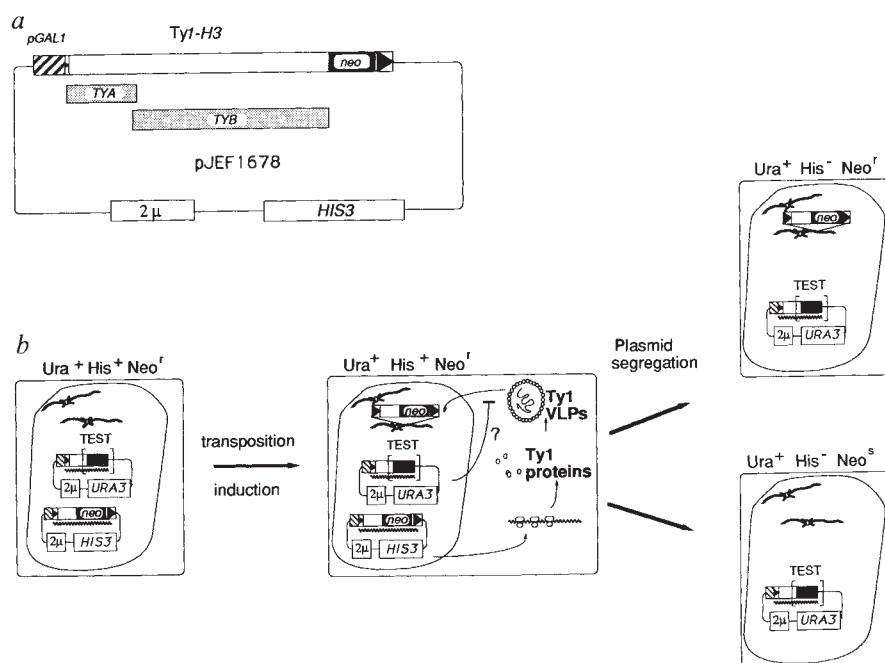
yeast cells. Because of this, mutants of BN were screened to identify those which were tolerated by yeast cells yet interfered with Ty1 transposition. The most desirable types of mutations would be those having an impact on specific activity of the nuclease, or on its ability to fold properly when not packaged in a VLP. Such mutant TYA-PR-BN proteins should be packaged into VLPs at a frequency relatively similar to that of native TYA-TYB proteins. Plasmid pGN1350.18 was mutagenized by hydroxylamine and transformed into yeast; of 15 large colonies obtained on glucose medium, six grew to some extent (some very well) on galactose.

Interference with Ty1 transposition was assayed by generating double transformants containing both a normal pGTy1 plasmid, pJEF1678, and either a mutagenized Ty1-BN plasmid, a Ty1-SN plasmid, or a control plasmid (Fig. 2b). The Ty1-SN fusions and control plasmids had no impact on Ty1 transposition, even in medium containing 100 mM CaCl₂ (Table 1). Presumably, the intracellular concentration of Ca²⁺ cannot activate SN, in accord with estimates of intracellular Ca²⁺ concentrations (micromolar range), far below the concentration required for SN activity (mM range)²⁵.

The mutant Ty1-BN plasmids allowing growth on galactose were also assayed for effects on transposition. Ty1 transposition frequency was inhibited by 99% in one of the six transformants (Table 1). This transformant contained a plasmid,

FIG. 2 A pGTy1 plasmid compatible with Ty1-nuclease plasmids for assaying *in vivo* and *in vitro* effects of potentially inhibitory plasmids. *a*, The structure of plasmid pJEF1678 is indicated. The plasmid is essentially identical to pGTyH3-*neo* (pJEF1105)²⁴ except that *URA3* has been replaced by *HIS3*. *b*, Method of assaying inhibition of Ty1 transposition using two plasmids.

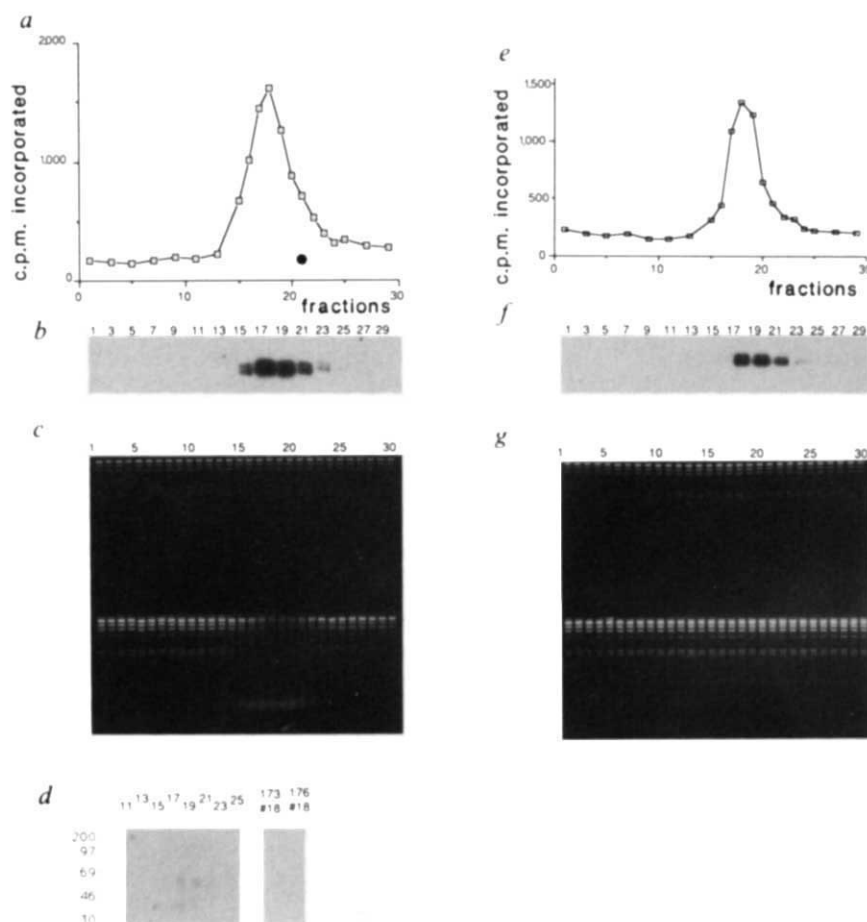
METHODS. *a*, Three DNA fragments were combined by ligation to generate pJEF1678. The *Bst*EII-*Bam*HI and *Sma*I-*Bst*EII fragments from pJEF1105 were combined with a *Bam*HI-*Sac*I fragment containing *HIS3* from plasmid pRS17 (R. Sikorski, personal communication). The *Sac*I end of the latter was rendered flush by treatment with Klenow fragment. *b*, The starting strain is a yeast transformant, harbouring pJEF1678. The *neo* marker in this plasmid confers resistance to the drug G418, an analogue of neomycin that is toxic to normal yeast cells. The phenotype of strains bearing pJEF1678 is His⁺Neo^r. The strain may or may not contain a second, potentially inhibitory plasmid, bearing the *URA3* marker. The *GAL1* promoters are induced by growth on galactose-containing medium. Ty1-*neo* messenger RNA is translated; Ty1 proteins assemble to form a virus-like particle that packages the Ty1-*neo* RNA. Inhibitory fusion proteins bearing TYA sequence at their N termini may coassemble into these VLPs. If Ty1 RNA remains intact, the TYB-encoded reverse transcriptase converts Ty1-*neo* RNA into double-stranded cDNA, which is integrated into the host



genome. The plasmid is then allowed to segregate by growth on histidine-containing medium. A few hundred cells are then plated on rich medium and allowed to form colonies. These colonies are replica-plated to a medium lacking histidine (SC-his) and to YPD medium containing the drug G418. The transposition fraction is the portion of total His⁻ colonies scored that are G418 resistant (Table 1).

FIG. 3 Assembly into Ty1 VLPs and nuclease activity of TYA-PR-SN fusion protein. Ty1 VLPs were prepared from strains GN173, containing the TYA-PR-SN fusion plasmid pGN1306 (*a-d*), and GN176, containing control vector plasmid pCGS109 (ref. 29) (*e-g*). Both strains also contain plasmid pJEF1678 (Fig. 2*a*). The strains were grown under conditions that induce high levels of transposition and VLP formation. Partially purified VLPs were then subjected to sedimentation on a linear 15–50% sucrose gradient. The gradients were fractionated and analysed biochemically; numbers indicate fraction. *a, e*, Reverse transcriptase activity; *b, f*, immunoblot using anti-TYA antibody; *c, g*, Ca²⁺-dependent nuclease activity; the upper set of wells in each panel represents reactions done without Ca²⁺, the lower in its presence. *d*, Immunoblot of selected fractions with anti-SN antibody.

METHODS. Ty1 VLPs were purified on step gradients as described²². The peak of RT activity was pooled and loaded on a second 15–50% linear sucrose gradient. Thirty 1.2 ml fractions of this gradient were collected and assayed for RT activity (*a, e*), anti-TYA immunoreactivity (*b, f*), nuclease activity (*c, g*) and anti-SN immunoreactivity (*d*). The left part of the Fig. (*a-d*) contains the results for GN173 (nuclease-containing experimental strain) and the right part (*e-g*) compares the results with GN176 (control strain). The RT assay was performed as described¹⁷. The SN enzymatic assay was performed as follows: 2 μl of each fraction was incubated for 30 min at 30 °C in a 20 μl solution containing 0.5 μg λ DNA restricted with *Hind*III in Hepes-KOH, 50 mM, pH 7.8, with or without CaCl₂ (10 mM). The product of this reaction was electrophoretically separated on 0.8% agarose. The immunoreactivity against TYA and SN proteins was assayed by immunoblotting using ¹²⁵I-labelled protein A to detect immune complexes.



pGN1350.18.7, that had no adverse impact on yeast growth on galactose or glucose media. The transposition-inhibiting effect observed is due to pGN1350.18.7 and specifically to its BN gene segment. The mutant plasmid had the expected structure; the BN fragment was sequenced. A new mutation resulting in substitution of a Phe codon for the Leu codon at position 89 was observed. The BN-K27A, L89F fragment was recloned and inhibited 98% of transposition (Table 1).

These results suggest that fusions between virion proteins and nucleases or other destructive enzymes could be useful in inhibiting viral infections in a variety of organisms, including man. Nucleases to be used in therapeutic CTVI should have minimal impact on host cells. Because SN interferes with neither cell growth nor Ty1 transposition, a noninfectious, intracellular pro-

cess¹⁷, SN may represent the best choice for combating viral infections in mammals. Intracellular Ca²⁺ levels are low, resulting in inactive intracellular capsid-nuclease fusion proteins. But when viruses containing these fusion proteins are shed into extracellular fluids, such as the bloodstream, encapsidated nucleases should be activated by the mM Ca²⁺ concentration. Further screening of TYA-PR-BN mutants and double mutants may identify additional BN mutants that are only active in the biochemical environment of VLPs or virions and inactive before assembly. We have described nucleases as the destructive moiety of fusion proteins for CTVI. But the use of other destructive enzymes, such as proteases and lipases, can also be envisioned to selectively destroy virion proteins and envelopes following appropriate targeting. □

Received 11 March; accepted 14 June 1991.

- Herskowitz, I. *Nature* **329**, 219–222 (1987).
- Baltimore, D. *Nature* **335**, 395–396 (1988).
- Boeke, J. D. in *Mobile DNA* (eds Berg, D. E. & Howe, M. M.) 335–374 (American Society for Microbiology, Washington DC, 1989).
- Atkins, J. F., Weiss, R. B. & Gesteland, R. F. *Cell* **62**, 413–423 (1990).
- Yoshinaka, Y., Katoh, I., Copeland, T. D. & Oroszlan, S. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1618–1622 (1985).
- Yoshioka, K. *et al.* *EMBO J.* **9**, 535–541 (1990).
- Brierley, C. & Flavell, A. J. *Nucleic Acids Res.* **18**, 2947–2951 (1990).
- Jones, T. A., Blaug, G., Hansen, M. & Barklis, E. J. *Viral.* **64**, 2265–2279 (1990).
- Weldon, R. A., Erdie, C. R., Oliver, M. G. & Wills, J. W. *J. Virol.* **64**, 4169–4179 (1990).
- Adams, S. E., Dawson, K. M., Gull, K., Kingsman, S. M. & Kingsman, A. J. *Nature* **329**, 68–70 (1987).
- Hartley, R. W. *Trends biochem. Sci.* **14**, 450–454 (1989).
- Mariani, C., deBeuckeleer, M., Truettner, J., Leemans, J. & Goldberg, R. B. *Nature* **347**, 737–741 (1990).
- Tucker, P. W., Hazen, E. E. & Cotton, F. A. *Molec. cell. Biochem.* **22**, 67–77 (1978).
- Weber, D. J., Serpersu, E. H., Shortle, D. & Mildvan, A. S. *Biochemistry* **29**, 8632–8642 (1990).
- Mossakowska, D. E., Nyberg, K. & Fersht, A. R. *Biochemistry* **28**, 3843–3850 (1989).
- Boeke, J. D., Garfinkel, D. J., Styles, C. A. & Fink, G. R. *Cell* **40**, 491–500 (1985).

- Garfinkel, D. J., Boeke, J. D. & Fink, G. R. *Cell* **42**, 507–517 (1985).
- Mellor, J. *et al.* *Nature* **318**, 583–586 (1985).
- Müller, F., Brühl, K. H., Freidel, K., Kowalik, K. V. & Ciriacy, M. *Molec. Gen. Genet.* **207**, 421–429 (1987).
- Adams, S. E. *et al.* *Cell* **49**, 111–119 (1987).
- Youngren, S. D., Boeke, J. D., Sanders, N. J. & Garfinkel, D. J. *Molec. cell. Biol.* **8**, 1421–1431 (1988).
- Eichinger, D. J. & Boeke, J. D. *Cell* **54**, 955–966 (1988).
- Eichinger, D. J. & Boeke, J. D. *Genes Dev.* **4**, 324–330 (1990).
- Boeke, J. D., Xu, H. & Fink, G. R. *Science* **239**, 280–282 (1988).
- Cuatrecasas, P., Fuchs, S. & Anfinsen, C. B. *J. Biol. Chem.* **242**, 1541–1547 (1967).
- Sikorski, R. S. & Hieter, P. *Genetics* **122**, 19–27 (1989).
- Boeke, J. D., Eichinger, D. J., Castrillon, D. & Fink, G. R. *Molec. cell. Biol.* **8**, 1432–1442 (1988).
- Hartley, R. W. *J. molec. Biol.* **202**, 913–915 (1988).
- Rose, M. D. & Fink, G. R. *Cell* **48**, 1047–1060 (1987).

ACKNOWLEDGEMENTS. We thank D. Shortle for gifts of reagents and for discussions, R. W. Hartley and A. R. Fersht for BN plasmids and S. Desiderio, A. Gabriel and H. Tomb for helpful comments on the manuscript. Supported in part by the NIH to J.D.B. J.D.B. is supported in part by an American Cancer Society Faculty Research Award and G.N. is supported in part by a Merck Postdoctoral Fellowship.

Transformation suppressor activity of a Jun transcription factor lacking its activation domain

Alison Lloyd, Nina Yancheva & Bohdan Wasyluk*

LGME-CNRS, U184-INSERM, Institut de Chimie Biologique, Faculté de Médecine, 11 rue Humann, 67085 Strasbourg Cedex, France

THE oncoprotein c-Jun is thought to be a mediator of *ras* transformation as both its synthesis and activity as a transcription factor are stimulated by *ras* expression^{1,2}. But c-Jun co-operates with *ras* in transformation assays³, suggesting that they act along different pathways (reviewed in ref. 4). Here we show by means of a dominant-negative mutated transcription factor that c-Jun potentially in conjunction with other factors that interact with it is necessary for transformation by *ras*. The mutant Jun lacks an activation domain and blocks stimulation of transcription by several oncoproteins, including Ras, v-Src, polyoma middle T, c-Jun and c-Fos, as well as by the tumour promoter 12-O-tetradecanoylphorbol-13-acetate (TPA). The inhibition is specific for motifs that bind Jun: activation of an NF- κ B/Rel motif is not affected. This Jun mutant acts as an anti-oncogene in *ras*-transformed cells, generating non-transformed revertants that have acquired anchorage and density-dependent growth, as well as reduced tumorigenicity *in vivo*. Mutants of other transcription factors designed to inhibit transformation will enable us to study their role in signal transduction.

We have investigated the function of Jun in *ras* transformation using a *trans*-dominant mutation in Jun. Amongst a number of mutated transcription factors, we found that $\Delta 9$, v-Jun deleted

of its activation domain (Fig. 1a; ref. 5), is particularly potent. $\Delta 9$ also contains the *Escherichia coli* Lex A repressor DNA-binding domain fused to its amino terminus to facilitate detection in cell extracts. In transfection assays in mouse fibroblasts, $\Delta 9$ expression does not activate transcription of reporters containing sequences that bind either Jun (pG1F4, Fig. 1b, c; pG1B4, Fig. 2a; ref. 6) or Lex A (ref. 5). But $\Delta 9$ efficiently inhibits transactivation by c-Jun, c-Fos, or their combination

TABLE 1 Mutant v-Jun expression inhibits anchorage independent growth and tumorigenicity in nude mice

Cells	Colony formation soft agar		Tumours in nude mice	[³ H]thymidine uptake ($\times 10^3$)
	Number of colonies	Colony size		
NIH3T3	NI	NI	±	13 ± 1
DT	79	>500	++/+++	94 ± 4
DTneo	80	>500	NI	102 ± 4
DTR1	20	<100	±	54 ± 3
DTR2	5	5–7	±	43 ± 4
DTR3	8	5–10	±	98 ± 2
DTR4	3	3–5	±	NI
DTR5	11	5–7	±	NI
DTR6	15	1–3	+	108 ± 6

Soft agar growth assay: 10^4 – 10^5 cells in 4 ml agar (Noble, Difco; 0.3% in growth medium with 10% FCS) were poured onto a 4-ml base layer (0.5% agar in medium) in 60-mm plates. The cells were fed after one week, and colonies counted in 20 random fields after 10–14 days. The number of colonies is presented per 100 plated cells. Tumorigenicity in nude mice: 10^6 cells in 0.1 ml PBS were injected subcutaneously into nude mice (at least two for each clone). The results are given after 14 days. +++, Animal with tumour died; ++, tumour >3 cm in diameter; +, tumour 1–2 cm; ±, tumour <1 cm or no tumour; NI, not investigated. Growth in low serum: 3×10^4 cells in 1 ml DMEM + 0.5% FCS (24-well dish) were preincubated for 24 h and labelled with [³H]thymidine (0.5 μ Ci m^{-1}) for 24 h. DNA was extracted and counted.

* To whom correspondence should be addressed.

through Jun binding sites, either in the absence of other motifs (pG1F4, Fig. 1b, c) or combined with the ets motif, which binds the Ets oncoprotein, of the oncogene-responsive unit of the polyoma virus enhancer (pG1B4, Fig. 2a). $\Delta 9$ also inhibits activation of the oncogene-responsive unit by the expression of Ras, v-Src and polyoma middle T (Fig. 2a), and by the tumour promoter TPA. ($\Delta 9$ does not inhibit transcription of the co-transfected expression vectors; results not shown.) The NF- κ B/Rel motif mediates transcription activation by TPA, Ras and polyoma virus middle T, but does not bind Jun (reviewed in ref. 7). $\Delta 9$ does not inhibit TPA activation of the NF- κ B element (Fig. 2b), showing that it is specific for elements that bind Jun.

If c-Jun is a mediator of *ras* transformation, then $\Delta 9$ could act as an anti-oncogene in *ras*-transformed cells. We used DT

cells, Ki-*ras*-transformed NIH 3T3, because of their low rate of spontaneous reversion⁸. They were co-transfected with expression vectors for $\Delta 9$ and aminoglycoside 3'-phosphotransferase, and colonies were selected with G418. We obtained with $\Delta 9$ about 200 colonies per μ g selectable marker, whereas in the absence of $\Delta 9$ there were about 1,000 colonies per μ g. Adherent clones were selected by blowing medium on the cells from a pipette, a simple procedure that does not impose additional selective pressure. Fifty clones per μ g selectable marker were obtained with $\Delta 9$. By contrast, there were no adherent clones when the $\Delta 9$ expression vector was omitted, or with an expression vector for a fusion protein between full-length v-Jun and Lex A (ref. 5). JunB, which inhibits transformation by c-Jun⁹, also did not give adherent clones. K-*rev-1*, which was isolated because of its ability to revert these cells⁸, gave 40

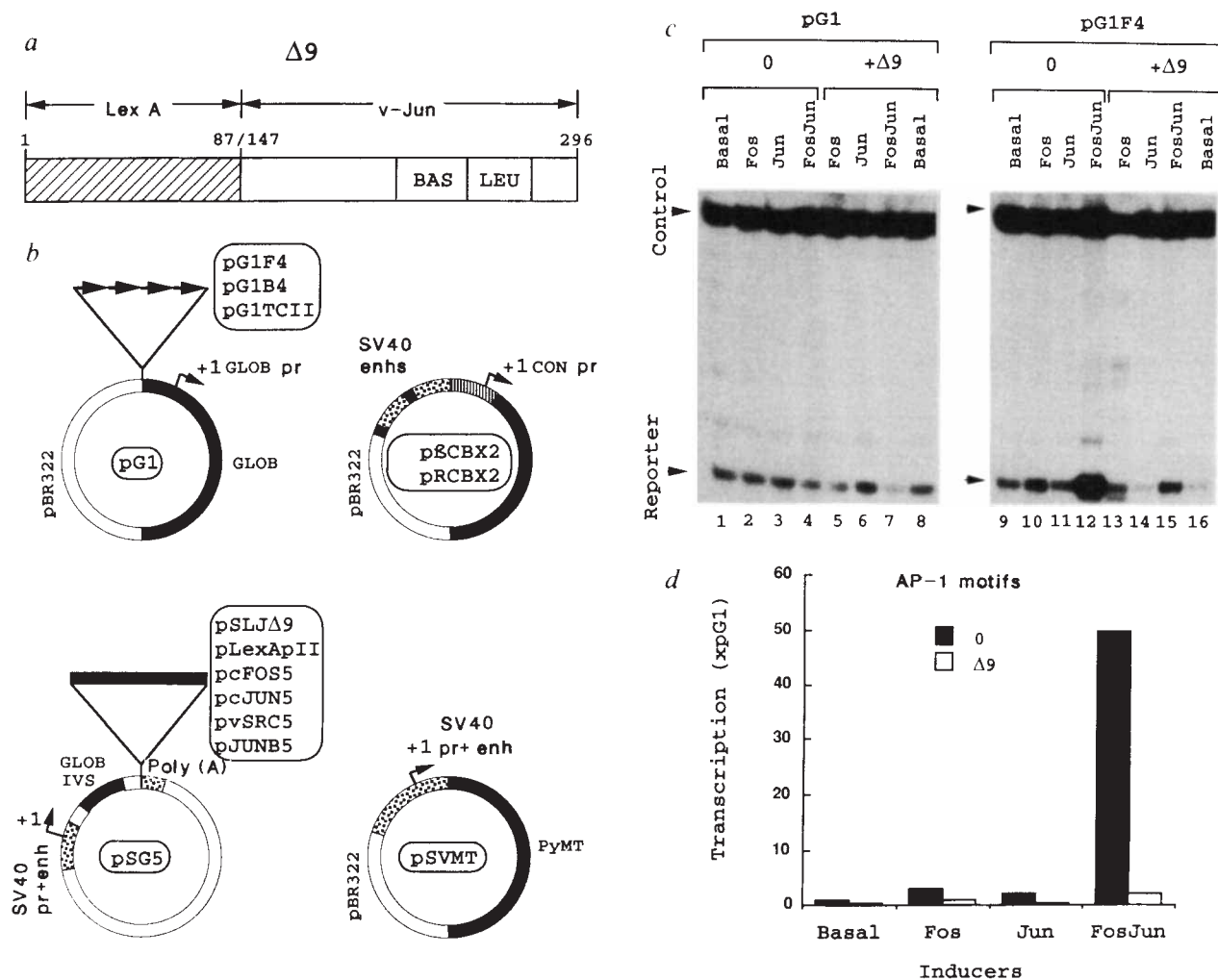


FIG. 1 Transdominant repression of c-Jun and c-Fos by a v-Jun derivative. **a**, Structure of the v-Jun mutant $\Delta 9$ (ref. 5) is a fusion protein between the DNA-binding domains of the oncoprotein v-Jun (amino acids 147-296, encompassing the basic region (BAS) and the leucine repeat (LEU)) and the bacterial repressor Lex A (amino acids 1-87). **b**, Reporters (top) and expression vectors (bottom). pG1 contains the rabbit β -globin gene (GLOB) including promoter sequences (GLOB pr) to -110. pG1F4, pG1B4 and pG1TCII contain four sequences with either the AP-1 motif (TGACTAA), the polyoma oncogene-responsive unit (GAGGAAGTGACTAA) or the Rel/NF κ B motif (TGGAAGTCCCCAG), respectively. The internal control p β CB $\times$ 2 has rabbit β -globin gene coding sequences (GLOB) downstream from the conalbumin promoter (CON pr) and two copies of the SV40 enhancer (SV40 enhs). The *ras* expression vector pRCB $\times$ 2 is similar to p β CB $\times$ 2, except that it has activated Ha-*ras* (RAS) gene coding sequences. The expression vector pSG5 contains the SV40 early promoter (SV40 pr+enh), β -globin intron II sequences and splice junctions (GLOB IVS II) and the SV40 poly(A) addition sequence. The derived vectors express $\Delta 9$ (pSLJ $\Delta 9$), the Lex A DNA-binding domain

(1-87, pLexApII), mouse c-Fos (pcFOS5), human c-Jun (pcJUN5), v-Src (pvSRC5) and mouse Jun-B (pJUNB5). In pSVMT, the SV40 early promoter drives transcription of polyoma middle T-coding sequences (PyMT). **c**, Expression of $\Delta 9$ inhibits c-Fos and c-Jun. Mouse fibroblasts (LMTK⁻) were co-transfected with reporter recombinants (1 μ g pG1F4 (lanes 1-8), or pG1 without the AP-1 motif multimer (lanes 9-16)), expression vectors (for c-Fos (1 μ g pcFOS5; lanes 2, 4, 5, 7, 10, 12, 13, 15), c-Jun (1 μ g pcJUN5; lanes 3, 4, 6, 7, 11, 12, 14, 15), and $\Delta 9$ (4 μ g pSLJ $\Delta 9$; lanes 5-8, 13-16) or the LexA DNA-binding domain (4 μ g pLexApII; lanes 1-4, 9-12)) and an internal control to correct for variations in transfection efficiency (0.5 μ g p β CB $\times$ 2; all lanes). After incubation in low serum (0.05%) for 36 h, RNA was extracted and analysed by quantitative S1-nuclease analysis for specific RNA initiated from the reporter and internal control recombinants (bands labelled REPORTER and CONTROL respectively) (see ref. 6). Electrophoresis was in 8% acrylamide-urea gels. **d**, Histogram of results in **c**. Transcription of pG1F4 is corrected against pG1. Black bars, Lex A DNA-binding domain; open bars, $\Delta 9$.

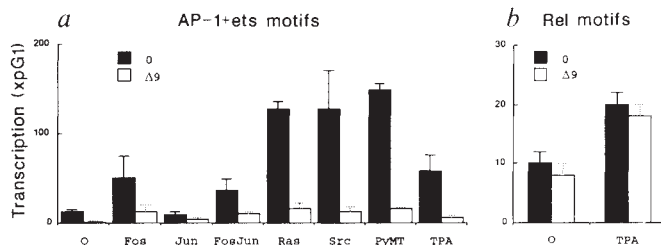


FIG. 2 Repression by $\Delta 9$ is specific for oncogene responsive elements containing AP-1 motifs. Mouse fibroblasts were transfected as described in Fig. 1 with reporters containing oncogene-responsive units that bind either AP-1 + ets (a, reporter pG1B4; see Fig. 1b) or Rel (b, reporter pG1TCII, see Fig. 1b) and appropriate expression vectors (Fos: pcFOS5; Jun: pcJUN5; Ras: pRCB $\times 2$; Src: pvsRC5; PyMT: pSVMT; O: pLexApII; $\Delta 9$: pSLJ $\Delta 9$; see Fig. 1b). TPA data: 200 ng ml $^{-1}$ TPA was added 24h before harvesting. Transcription is corrected against pG1. Black bars, Lex A DNA-binding domain; open bars, $\Delta 9$.

adherent clones per 1,000 G418-resistant clones and per μ g selectable marker. These results show that $\Delta 9$ efficiently and specifically generates adherent clones.

Six clones were chosen at random, isolated and grown into mass culture for further analysis. They all express the $\Delta 9$ protein (Fig. 3b). The levels of Ras protein (Fig. 3c) and RNA (results not shown) were similar to those from DT cells, and higher than in NIH 3T3 cells, demonstrating that decreased Ras expression could not account for their phenotype. They were reverted to various extents, depending on the criteria used: (1) The cells were flatter than either DT or DTneo cells (Fig. 3a) and showed

contact-inhibited growth; (2) they had lower levels of cathepsin L RNA (Fig. 3d), which is increased by *ras* transformation¹⁰ (Fig. 3d); (3) revertants form fewer and much smaller clones in soft agar (Table 1); (4) the clones form smaller, slower-growing tumours in nude mice (Table 1). Tumours from the revertant clones did not express detectable levels of $\Delta 9$ (results not shown), indicating that the loss of $\Delta 9$ expression (in the absence of selection medium) is a requirement for tumour formation. By contrast, the revertant clones grew relatively well in low serum compared with NIH 3T3 cells. This suggests that there is an independent *ras* pathway which does not involve Jun. Interestingly, both antisense and antibody-injection experiments show that Fos is required for DNA synthesis and growth in low serum in transformed cells¹¹⁻¹³, but there is no direct evidence for a transforming activity of Fos that does not involve heterodimer formation with Jun¹⁴.

$\Delta 9$ could block transformation by dimer formation with its leucine repeats, or by inhibiting binding of factors to Jun motifs, or by a combination of these mechanisms. Similar mechanisms have been evoked for Δ FosB, a product of the *fosB* gene which regulates Fos and Jun: like $\Delta 9$, Δ FosB retains its leucine repeat and basic region, but lacks an activation domain¹⁵. By contrast, Id, a protein related to myoD, lacks a basic region and inhibits myoD and muscle cell differentiation by forming heterodimers that do not bind to DNA¹⁶.

Oncogene co-operation is little understood. Nuclear and cytoplasmic oncogenes co-operate, which has led to the suggestion that they are on separate signalling pathways⁴. But many of the nuclear oncogenes are stimulated by cytoplasmic oncogenes, which is suggestive of a linear progression⁴. Here we have shown that c-Jun activity is required for the transforming effects of *ras*. As additional c-Jun is necessary for full transformation of

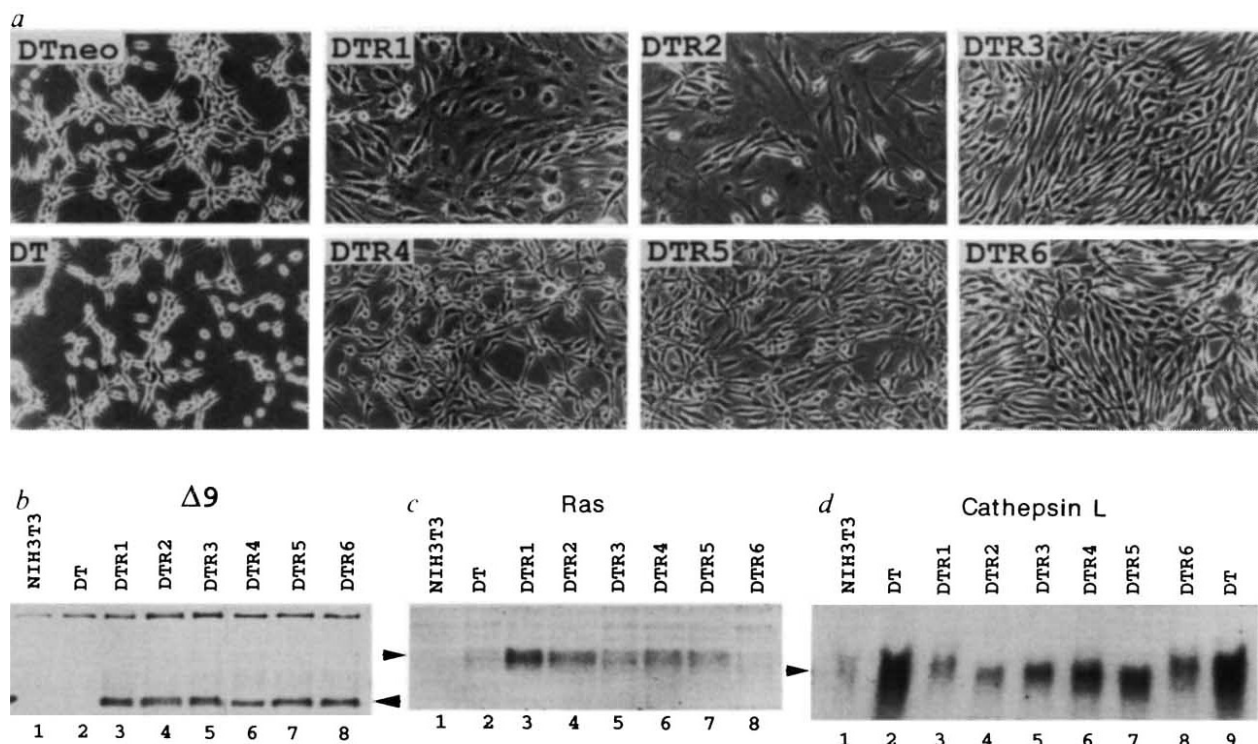


FIG. 3 The v-Jun mutant reverts Ki-Ras-transformed cells. DT cells¹⁷ were transfected using calcium phosphate with expression vectors for aminoglycoside 3'-phosphotransferase (pSV2neo; 1 μ g), and either $\Delta 9$ (pSLJ $\Delta 9$), K-rev-1 (ref. 17), Jun-B (pJUNB5), or the Lex A DNA-binding domain (pLEXApII) (9 μ g made up to 25 μ g per 9 cm plate with pSG5; 5, 6, 8, where 5, 6 and 8 refer to structures of expression vectors). Sixteen hours later the cells were washed, and the following day were trypsinized, diluted 10-, 20- and

100-fold and selected in 0.5 mg ml $^{-1}$ G418. After 9 d, colonies were counted and revertants selected by adhesion enrichment and isolated with cloning rings. Six clones (DTR1-6) were grown into mass culture and photographed with a phase contrast microscope (a); protein or RNA was isolated for analysis by blotting for expression of $\Delta 9$ (b, using rabbit anti-Lex A affinity-purified antibodies), Ras (c, using Y-13 259, rat monoclonal antibodies) and for cathepsin L (d, RNA blot with mouse cathepsin L probe).

primary cells³, oncogene co-operation may involve the unblocking of two regulating steps on the same signalling pathway. It should be possible to establish the role of other transcription factors (Fos, Ets, Myb, Rel, Myc, SRF; see ref. 7) in signal transduction and the transformed phenotype using mutants similar to $\Delta 9$. These reagents have the particular advantage that they are downstream in the pathways, having the potential to cause a wide range of transformed cells to revert. $\square$

Received 17 April; accepted 8 July 1991.

1. Imler, J. L., Schatz, C., Wasylyk, C., Chatton, B. & Wasylyk, B. *Nature* **332**, 275–278 (1988).
2. Sistonen, L., Hölttä, E., Mäkelä, T. P., Keski-Oja, J. & Alitalo, K. *EMBO J.* **8**, 815–822 (1989).
3. Schutte, J., Minna, J. D. & Birrer, M. J. *Proc. natn. acad. Sci. U.S.A.* **86**, 2257–2261 (1989).
4. Hunter, T. *Cell* **64**, 249–270 (1991).
5. Schneikert, J., Imler, J. L. & Wasylyk, B. *Nucleic Acids Res.* **19**, 783–787 (1991).
6. Wasylyk, C., Flores, P., Gutman, A. & Wasylyk, B. *EMBO J.* **8**, 3371–3378 (1989).
7. Gutman, A. & Wasylyk, B. *Trends Genet. Sci.* **7**, 49–54 (1991).
8. Kitayama, H., Sugimoto, Y., Matsuzaki, T., Ikawa, Y. & Noda, M. *Cell* **56**, 77–84 (1989).
9. Schutte, J. *et al. Cell* **59**, 987–997 (1989).
10. Kane, S. E. & Gottesman, M. M. *Cancer Biology* **1**, 127–136 (1990).
11. Ledwith, B. J., Manam, S., Kraynak, A. R., Nichols, W. W. & Bradley, M. O. *Molec. cell. Biol.* **10**, 1545–1555 (1990).
12. Mercola, D., Westwick, J., Rundell, A. Y. K., Adamson, E. D. & Edwards, S. A. *Gene* **77**, 253–265 (1988).
13. Riabowol, K. T., Vosatka, R. J., Ziff, E. B., Lamb, N. J. & Feramisco, J. R. *Mol. Cell. Biol.* **8**, 1670–1676 (1988).
14. Schuermann, M. *et al. Cell* **56**, 507–516 (1989).
15. Nakabeppu, Y. & Nathans, D. *Cell* (1991) **64**, 751–759 (1991).
16. Benezra, R., Davis, R. L., Lockshon, D., Turner, D. L. & Weisraub, H. *Cell* **61**, 49–59 (1990).
17. Noda, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **86**, 162–166 (1989).

ACKNOWLEDGEMENTS. We thank C. Marshall for reagents, C. Benoist, F. Bibollet and J. M. Limacher for help with the nude mice, members of the laboratory for discussion, and the cell culture, oligonucleotide, animal house, secretarial and photographic services staff for their help, and CNRS, INSERM, ARC and FNCLCC for financial assistance. A. L. was a recipient of EMBO and EEC cancer training fellowships.

Characterization of the rough endoplasmic reticulum ribosome-binding activity

Jodi M. Nunnari, Deborah L. Zimmerman, Stephen C. Ogg & Peter Walter

Department of Biochemistry and Biophysics, University of California Medical School, San Francisco, California 94143-0448, USA

THE rough endoplasmic reticulum membranes of mammalian cells contain specific ribosome-binding sites¹. A purification to apparent homogeneity of a negatively charged protein (ERp180) of relative molecular mass 180,000 (180 K) was reported which was proposed to function as a rough endoplasmic reticulum ribosome receptor². We report here that ribosome-binding site activity quantitatively solubilized from rough endoplasmic reticulum membranes does not cofractionate with ERp180. By contrast, ribosome-binding site activity fractionates as a much smaller, positively charged protein.

The density of ribosome-binding sites present in canine pancreatic rough microsomal membranes (RM) was determined after dissociation of ribosomes from RM with high salt after the nascent proteins were released with puromycin. Ribosome-binding sites on the stripped RM were quantitated by rebinding [¹²⁵I]-labelled ribosomes¹. Scatchard analysis indicated that the density of sites on the RM was 350 fmol per equivalent (or 120 nmol per g membrane protein; not shown), in good agreement with previous estimates³.

Ribosome-binding sites were solubilized using a two-step procedure modified from Yoshida *et al.*⁴. Endoplasmic reticulum luminal proteins were released by lysing RM in a nonionic detergent in 50 mM KCl. Membrane proteins were then solubilized using the detergent CHAPS under high-salt conditions. The ribosome-binding activity of the extracts was assayed after reconstituting the membrane proteins into liposomes. Figure 1

shows that the number of binding sites recovered increased as the ratio of CHAPS to protein was increased during solubilization. At a ratio of detergent to protein of 12.5 μ g CHAPS per mg protein or greater, a maximum yield of 320 fmol ribosome-binding sites per equivalent of RM was obtained. Hence, more than 90% of the sites present in RM were solubilized and functionally recovered after reconstitution into liposomes. The extraction and reconstitution procedure allowed us, therefore, to account quantitatively for the ribosome-binding sites present in the starting RM fraction.

We fractionated the CHAPS extract by velocity sedimentation (Fig. 2) or by ion-exchange chromatography (Fig. 3). The peak of ribosome-binding activity sedimented near the top of the sucrose gradient (<4S) (Fig. 2a). Most (80%) of the sites did not bind to the DEAE resin (Fig. 3a). The fractionation on the DEAE column resulted in a large increase in the specific activity in the flow-through fraction and a concomitant decrease in the eluate fraction (Fig. 3a). Scatchard analysis of ribosome-binding activity after DEAE fractionation indicated a dissociation constant (K_d) of 12 nM, experimentally indistinguishable from ribosome binding to RM (data not shown; ref. 3). In addition, the salt and protease sensitivity of ribosome binding to these proteoliposomes and RM fractions were indistinguishable (data not shown). Thus, the ribosome-binding activity recovered in the flow-through fraction of the DEAE column had all the properties characteristic of that of the starting RM fraction.

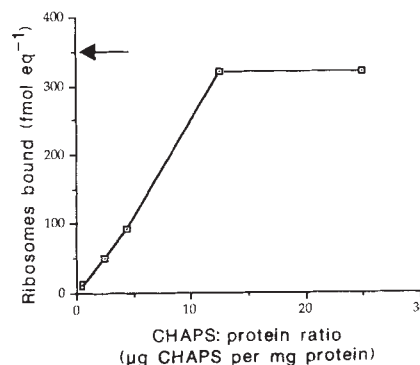
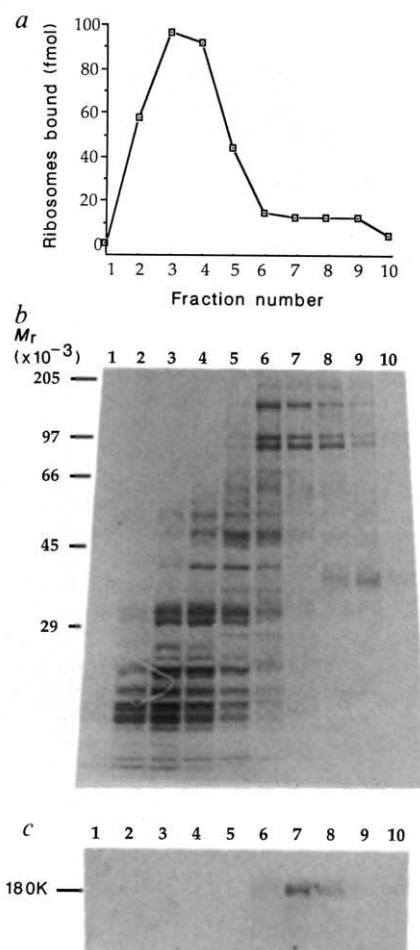


FIG. 1 Quantitative solubilization of ribosome-binding activity. A CHAPS extract of RM was prepared at different detergent-to-protein ratios. The ribosome-binding activity in the extracts was determined after reconstitution into liposomes. The arrow indicates the binding activity present in an equivalent amount of ribosome-stripped RM.

METHODS. Dog pancreatic endoplasmic reticulum membranes (1 eq μ l⁻¹; defined in ref. 8), isolated as described previously⁸, were extracted initially using 1.5% Emulgen-913 and 50 mM KCl in buffer A containing 50 mM Tris, pH 7.5, 1 mM dithiothreitol, 0.5 mM phenylmethylsulphonyl fluoride, 1.0 μ g ml⁻¹ each of antipain, leupeptin, chymostatin and pepstatin for 30 min at 0 °C. Membranes were pelleted through a 0.6 M sucrose cushion in buffer A containing 1.5% Emulgen-913 and 50 mM KCl in a Beckman 50.2 Ti rotor at 40,000 r.p.m. for 2 h at 4 °C. Pellets were resuspended by homogenization at final protein concentrations (Bio-rad protein assay) varying from 0.8 to 40 mg ml⁻¹ (0.1–5.0 eq μ l⁻¹) in buffer A containing 2% (w/v) CHAPS, 10% sucrose and 500 mM KCl and allowed to equilibrate at 0 °C for 30 min. CHAPS extracts were centrifuged in a 50.2 Ti rotor at 50,000 r.p.m. for 3 h at 4 °C. Supernatants were adjusted with buffer A containing 2% CHAPS, 10% sucrose and 500 mM KCl to a final concentration of 0.1 eq μ l⁻¹. To 100 μ l (10 eq) of each of the supernatants, 0.5 mg of egg phosphatidylcholine, solubilized in 50 mM Tris, pH 7.5, 50 mM KCl, 0.25 mg sodium cholate and 0.1 mg CHAPS, was added and reconstituted by dialysis against 1,000 volumes of 50 mM Tris, pH 7.5, and 50 mM KCl for 16–40 h with one change of the dialysis buffer. Proteoliposomes (50 μ l, 5 eq) were assayed for ribosome binding as previously described¹ with the modification that [¹²⁵I]-labelled rat liver ribosomes at a concentration of 50 nM were used as the substrate. Rat liver ribosomes were prepared as described¹ and iodinated using Bolton-Hunter reagent for 1 h at 0 °C and subsequently purified on a 10–30% sucrose gradient containing 50 mM Tris, pH 7.5, 50 mM KCl and 5 mM MgCl₂.

FIG. 2 Sucrose gradient centrifugation of CHAPS-solubilized extract. A CHAPS extract of RM was prepared at the optimal detergent to protein ratio (Fig. 1) and fractionated by sucrose gradient centrifugation. Sedimentation was from left to right. *a*, Fractions were analysed for ribosome binding after reconstitution into liposomes (Fig. 1). The recovery of ribosome-binding activity after fractionation approached 100%. *b*, *c*, Fractions were analysed by SDS-PAGE and staining with Coomassie blue (*b*) or immunoblotting with anti-180K serum (*c*).

METHODS. CHAPS extract was prepared as described in Fig. 1 at a protein-to-CHAPS ratio of 12.5 μg CHAPS per mg protein. CHAPS extract (200 μl , 40 eq) was layered on a 2.0 ml 5–20% sucrose gradient in buffer A containing 0.5% CHAPS and 500 mM KCl and centrifuged in a Beckman TLS55 rotor at 55,000 r.p.m. for 12 h at 4 °C. Gradients were fractionated into 200 μl aliquots. Each fraction (100 μl) was analysed for ribosome-binding activity as described in Fig. 1, with the exception that 4 nM ^{125}I -labelled ribosomes were used in the assay. Separate 100 μl aliquots were analysed by SDS-PAGE, followed by either Coomassie-blue staining or immunoblotting for the 180K protein. For western blotting, proteoliposomes were subjected to SDS-PAGE and transferred onto nitrocellulose. Blots were incubated with rabbit antiserum raised by injection of SDS-PAGE purified ERp180 (D.L.Z. and P.W., unpublished results). Immune complexes were visualized using ^{125}I -labelled goat anti-rabbit IgG and autoradiography.



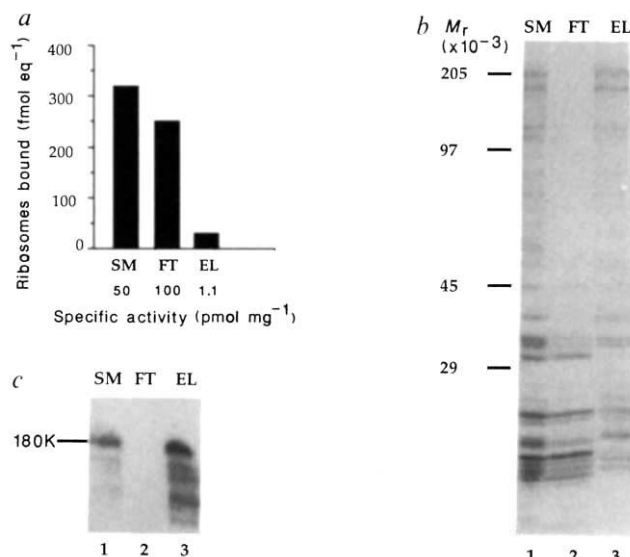
Savitz and Meyer reported the identification of ERp180 which they proposed to be a rough endoplasmic reticulum ribosome receptor². They identified a 160K proteolytic fragment which was able to inhibit ribosome binding to RM. Antibodies raised against this proteolytic fragment were used to purify the parent protein, ERp180. Both the sedimentation behaviour and the charge properties of ERp180 are different from the ribosome-

binding activity characterized in Figs 2 and 3.

To address this discrepancy, we used antibodies raised against ERp180 purified by SDS-PAGE. The antiserum recognized a 180K membrane protein and a 160K soluble fragment released from RM after treatment with thermolysin (data not shown), indicating that it was specific for ERp180. The antiserum was used to follow ERp180 during sucrose gradient sedimentation

FIG. 3 DEAE chromatography of CHAPS-solubilized extract. A CHAPS extract of RM (as in Fig. 2) was fractionated on DEAE-Sepharose. The initial extract (lane 1, 'SM'), a low-salt flow-through fraction (lane 2, 'FT'), and a high-salt eluate fraction (lane 3, 'EL') were collected. *a*, Equivalent amounts of each fraction were analysed for ribosome binding. The bar graph displays the total ribosome-binding activity recovered in the respective fractions; the specific activity for each fraction calculated as pmol ribosome-binding site per mg protein is indicated. *b*, *c*, Equivalent amounts of each fraction were analysed by SDS-PAGE and staining with Coomassie blue (10 eq per lane, *b*) or immunoblotting with anti-180K serum (5 eq per lane, *c*). During fractionation on the DEAE column some degradation of the 180K protein was observed in the eluate fraction.

METHODS. CHAPS extract (1 ml, 200 eq) was diluted 10 $\times$ and adjusted to 0.5% CHAPS, 50 mM KCl with buffer A and a 15% CHAPS stock solution. The extract was passed over a 2 ml DEAE-Sepharose Cl-6B column equilibrated with buffer A, containing 0.5% CHAPS, 50 mM KCl. The column was washed with three column volumes of the equilibrating buffer, which was combined with the flow-through fraction. The column was eluted with buffer A, containing 0.5% CHAPS, 500 mM NaCl. For saturation binding analysis, 500 μl of the DEAE flow-through fraction was reconstituted into liposomes by the addition of 2.5 mg egg phosphatidylcholine, solubilized in 50 mM Tris, pH 7.5, 50 mM KCl, 1.25 mg sodium cholate and 0.5 mg CHAPS and by dialysis as described in Fig. 1. Fractions (29 μl , 0.57 eq) were assayed for ribosome-binding activity using ribosome concentrations ranging from 10 to 100 nM as described in Fig. 1.



(Fig. 2c) or during the DEAE-Sepharose fractionation (Fig. 3c) of the CHAPS extract. Note that in both cases ERp180 did not cofractionate with the ribosome-binding activity.

These discrepancies could be due to differences between the solubilization and reconstitution procedures described here and those used by Savitz and Meyer, but given the nearly quantitative solubilization and reconstitution of the ribosome-binding activity we observed, this explanation is unlikely. Moreover, under our conditions ERp180 was incorporated into the proteoliposomes efficiently and in the proper orientation (data not shown).

Chemical crosslinking studies with endogenous membrane-bound ribosomes have demonstrated that some ERp180, in addition to other endoplasmic reticulum membrane proteins including ribophorins, could be found in proximity to membrane-bound ribosomes⁵. Ribophorins were previously proposed to function in ribosome binding⁶, but were subsequently shown not to be required^{3,4}. Recently, crosslinking of ribosomes to a 34K membrane protein was observed with reconstituted proteoliposomes which exhibited ribosome-binding activity⁷. Thus, on the basis of crosslinking studies the environment of the membrane-bound ribosome seems complex and, formally, the possibility remains that ERp180 contacts ribosomes. But the data presented here demonstrate that the ribosome-binding activity of RM, as biochemically defined by a quantitative binding assay, resides in a protein that still awaits purification. □

Received 13 May; accepted 2 July 1991.

1. Borgese, N., Mok, W., Kreibich, G. & Sabatini, D. D. *J. molec. Biol.* **88**, 559–580 (1974).
2. Savitz, A. J. & Meyer, D. I. *Nature* **346**, 540–544 (1990).
3. Hortsch, M., Avossa, D. & Meyer, D. I. *J. Cell Biol.* **103**, 241–253 (1986).
4. Yoshida, H. et al. *Biochem. J.* **259**, 811–819 (1987).
5. Collins, P. G. & Gilmore, R. *J. Cell Biol.* (in the press).
6. Kreibich, G., Ulrich, B. L. & Sabatini, D. D. *J. Cell Biol.* **77**, 464–487 (1978).
7. Tazawa, S. *J. Biochem.* **109**, 89–98 (1991).
8. Walter, P. & Blobel, G. *Meth. Enzym.* **96**, 682–691 (1983).

ACKNOWLEDGEMENTS. We thank C. Nicchitta, J. Miller, P. Collins and R. Gilmore for helpful discussions. This work was supported by the American Cancer Society (to J.M.N.), the Lucille P. Markey Charitable Trust (to D.L.Z.), the Howard Hughes Foundation (to S.C.O.) and the NIH and Alfred P. Sloan Foundation (to P.W.).

Multiple nucleotide-binding sites in the sequence of dynein β heavy chain

I. R. Gibbons, Barbara H. Gibbons, Gabor Mocz & David J. Asai*

Pacific Biomedical Research Center, University of Hawaii, Honolulu, Hawaii 96822, USA

* Department of Biological Science, Purdue University, West Lafayette, Indiana 47907, USA

AXONEMAL dyneins have two or three globular heads joined by flexible tails to a common base, with each head/tail unit consisting of a single heavy-chain polypeptide of relative molecular mass >400,000. The sizes of the components have been deduced by electron microscopy^{1–3}. The isolated β heavy chain of sea urchin sperm flagella, which is immunologically identical to that of the embryo cilia (data not shown; ref. 4), is of particular interest as it retains the capability for microtubule translocation *in vitro*^{5,6}. Limited proteolysis of the β heavy chain divides it into two fragments, A and B, which sediment separately at 12S and 6S, and possibly correspond to the head and tail domains of the molecule⁷. Dynein ATPase is the energy-transducing enzyme that generates the sliding movement between tubules that underlies the beating of cilia and flagella of eukaryotes, and possibly also other large

intracellular movements^{8,9}. Here we report that the deduced amino-acid sequence of the β heavy chain of axonemal dynein from embryos of the sea urchin *Triploneustes gratilla* has 4,466 residues and contains the consensus motifs for five nucleotide-binding sites. The probable hydrolytic ATP-binding site can be identified by its location close to or at the V1 site of vanadate-mediated photocleavage¹⁰. The general features of the map of photocleavage and proteolytic peptides reported earlier have been confirmed, except that the map's polarity is reversed. The predicted secondary structure of the β heavy chain consists of an α/β -type pattern along its whole length. The two longest regions of potential α helix, with unbroken heptad hydrophobic repeats 120 and 50 amino acids long, may be of functional importance. But dynein does not seem to contain an extended coiled-coil tail domain.

Progress in analysing the structure and function of dynein has been hindered by the lack of primary sequence information resulting from the difficulty of cloning a complementary DNA encoding such a large polypeptide. We have now used a polymerase chain reaction (PCR)-directed procedure to obtain the complete sequence of the β heavy chain of dynein from sea urchin embryos (Fig. 1). In the region of overlap our sequence shows about 98% identity to two partial clones of dynein β heavy chain from a different species of sea urchin examined by Ogawa^{12,13}, but no significant similarity to two other partial sequences^{14,15}.

The relative positions of the N-terminal microsequences of the tryptic peptides (relative molecular masses 130,000 (130K) and 124K) on the L13–14 parent clone and a short upstream extension of it indicated early that the polarity of the peptide chain must be reversed from what had been deduced previously by assay of acetyl residues in the photocleaved β chain¹¹. Further sequence data showing the positions of the other microsequenced peptides, as well as the probable position of the V1 vanadate-mediated photocleavage site¹⁰, unambiguously confirm the new polarity. A revised map is shown in Fig. 2.

A striking feature of the sequence is the presence of five copies of the consensus A motif¹⁶ of nucleotide-binding sites (A/G)XXXXGK(T/S) (single-letter amino-acid code; Figs 1 and 3a). The principal hydrolytic ATP-binding site seems to be the GPAGTGKT sequence (GKT2) at residues 1,852–1,859, for this position 660 amino acids downstream from the N terminus of the 215K tryptic peptide corresponds closely to the expected location of the V1 photocleavage site on the peptide map (Fig. 2). By analogy to the known location of vanadate-mediated photocleavage in myosin and in adenylate kinase^{17,18}, proline 1,853 in the proposed hydrolytic ATP-binding site of the β chain is a likely candidate for the site of scission in V1 photocleavage. The sequence around the hydrolytic ATP-binding site of the β chain does not closely resemble that of the other known cytoskeletal motor proteins; in particular, the flexible glycine-rich loop extends upstream beyond the GXXXXGKT motif and is longer than that in most other ATP-binding proteins (Fig. 3, and ref. 19).

A second GKT nucleotide-binding motif (GKT1) occurs at residues 154–161, and two GKS motifs are located at 2,133–2,140 and 2,460–2,467. A modified GKQ motif occurs at 2,805–2,812 (Figs 1 and 3a). Although the function of GKT2, the hydrolytic ATP-binding site at 1,852–1,859 is clear, the roles of the two GKS sites, the GKQ site and the GKT1 site are much less so. The almost identical sequence (GNAGXGKS) of the two GKS motifs (Fig. 3a) suggests that they constitute parts of two nucleotide-binding sites that have been conserved in evolution and represent an important functional aspect of the dynein β chain. Possible roles include the regulation of dynein crossbridge activity which is required for the oscillatory beating of cilia and flagella, or involvement in the inhibitory action of high ATP concentrations on axonemal dynein ATPase activity²⁰. The sizes of the peptides formed by cleavage at the unique site of Fe(III)-mediated photocleavage²¹ indicate that it lies close to the GKS2 site. Similarly, the roughly 100K mass of the peptide formed by double photocleavage at the V1 and V2 sites²² makes serine

FIG. 1. Deduced amino-acid sequence (in single-letter notation) of the β -heavy chain of outer arm axonemal dynein from sea urchin embryos, numbered from the putative start methionine. Residues that were confirmed by direct N-terminal sequencing of tryptic peptides of dynein β heavy chain are underlined. Overlined residues indicate locations of nucleotide-binding consensus sequences.

METHODS. Peptide sequencing: four major tryptic peptides of the dynein β chain from sea urchin sperm flagella were separated electrophoretically and sequenced on Immobilon-PDVF blots with an Applied Biosystems Model 470A sequencer. Starting with the second residue of the peptide in each case, the sequences obtained were: 110K, LLPLPVGSXKV; 215K, QVAPLQANEVAL; 130K, LAAGLXSWVNVIVKFNVY; 124K, ITLANXLVGG-LASEN. DNA sequencing: a randomly primed, size-selected cDNA library was constructed in λ gt11 using poly(A)⁺ RNA prepared from mesenchyme blastula embryos of the sea urchin *Tripneustes gratilla* that had been allowed to regenerate cilia for 60 min after deciliation by hypertonic shock. The library was amplified and subjected to immunoscreening and northern blot analysis for selection of candidate clones, from which the parent clone (L13-14) of 1,179 base pairs was chosen. Extension of the parent clone sequence was obtained by using successive steps of PCR-amplification to walk through the cDNA library. Two nested specific primers and a general primer based on the sequence of the vector were used at each step³⁴. Several lines of evidence support the identity and integrity of the sequence derived by extension of this clone: (1) L13-14 yields a fusion protein that reacts with affinity-purified polyclonal anti- β -heavy chain antibody; (2) the epitope-selected antibody obtained by adsorbing antibody with L13-14 fusion protein reacts specifically with the 130K peptide on immunoblots of tryptic digests of β heavy chain; (3) DNA probes prepared from L13-14, as well as from five other segments along the length of the cDNA sequence, all hybridize on northern blots to a poly(A)⁺ RNA of the large size (about 14.5 kilobases) expected for the β -chain mRNA; (4) directly obtained micro-sequence of the N termini of four tryptic peptides distributed along the length of the β heavy chain are all found in the deduced amino-acid sequence at positions consistent with the known arrangement of the tryptic peptides; (5) the predicted polypeptide of 4,466 amino acids has an M_r of 511,804K, close to the expected mass for the β heavy chain (about 475,000K); (6) the predicted polypeptide contains the consensus motif for an ATP-binding site at the position expected for the V1 photocleavage site.

```

10      20      30      40      50      60      70      80      90     100
1  MADVDPRLEFISEYILKSYKLPDKWAKCINVEENKILMFLFKADNPQLVFTVNAALITPSYFSPALKNTKATYIFIKKGRPVGDNIKTLLTVYG
101 DLSYTPLEQLSALVDEILVPLLANPRNHQMPVVSQDLRVHNLKSSVYVAGVQVXKTLPLPVGSEKVEATAESEEKDDSYDRSLVHAIESVIDW
201 THQTRDLVKRDSAPLLEGLNPGMVEINFUKAKCENLDCIFQQLRDPKVRMKELLERTQSSYLSFNMIRDEVAALTEADQINCYLPLIYQVESLD
301 ELEFPDMTPRLAPILHTVCLISWNSDYNTAPRIIVLLQECINLLIDLCRTFLDPSIEFKLEPEESLEKVRGALAVLKAWRDLYDEHRAKLDKYFKDGRE
401 VKGEWAFSLVFTRMNFTSRIETIQSLFETNVEFSKLEKTEGSMKGRMLSQQVEKIEEFQECACVFTERPVDGLDQCQEFLLDDEYFEKKVFDLDR
501 RLGSILCGFDCCGLEAVFKMLDCYGPLLDRPVRINDFECKYPVVMMLYDLEDOAKEIYDEHMRVEENDGNAPLNKMPDVGQLKWSAQLRDRISKP
601 MGSLSKMEHPTGVRRILESEDAKVYFQYKEMINLLNKYEKQVFENWTKGVDEVCXTNLQSLITRDESKLIKINFPDKPLVAVREVYTLQIRGEESIP
701 ESAASIEYKHETLRKYVANLDTQAWYNKVRNTVLEVEFPIEGQLADLTRLKQAESELNWTSQSWVEYIQTETRDGVHLEKRVQQTQDNVDRICKKIM
801 EWTQKPLFERKELKESLLALDRQDRGLKKRYAIEISTAGEKIHSLIKENGLFKADASSIWKAYVDYDDMDVDFGNCIHCTLSYLLENDPRHCAAP
901 LFEARLELQVPMIFNPSLDYGVADGFYDLVEMLSDTYKMASLSVSLAEHNGQEHYQADLEGMDLSDVRNLDMDRVGTIMTKAQEYRNSFDNYAYLYV
1001 DORKEFMRFLLYHNHVLTEETEAHAEDGVPCPTLDQFKEQVDTYEKIYSEADEIEPEQVDFAWFRVDSKPFKALNIIKKVSFMFKQLIDHVTNS
1101 LSELQEFIKVGNGLTKTVEEGDNGYGLVECMHLMVAKRQATDEMFEPIKQTEILLKTYDQEMSEEVHTQLQELPEQWNNTKKIATIVKQVAPLQAN
1201 EALIRRKCTSFDRQHEFRFRERFKEAPFITFEGPYCQLDRCHSEIYEMEEHMANLESAGLFEVMPDYKQKACRVERLLKALWIMIVRTSIED
1301 WKTPTWLEINVEQEMEDCKFKAKIRSLQKEMRAWDAYNGLDATVKNMNTSLRAVSELQNAIRERHWQQLMAATKYKFTMDKETTLLSALLNLHNFED
1401 EVRNIYDKAVKEMGEKVELNLTWSSMDFEYPSHRTGISLLKSNEELIETLEDNQVQLNMLSKHIAHFLEEYSGWQKLLSTDSVITIFVEVQRT
1501 WSHLESIFIGSEDIRNQLPEDSKFRGIDTDFKELASEMEKTPNVVEATNRARLYDRLEAIGSLVVECKALAEYLETKRLLAFPRFYFSSADLLILSQ
1601 GNPSPQVORHLSKFLDNMAKLFKQDDGENDKTALGMYSKEGEYVDFKCECTGQVEVLRNVRMDAMRSTVRSQFADAVVSEYKPREQVLYDPAQV
1701 ALATTQVWITTEVNIISFARLEEGHENSMDYKQKIQOQLNTLIGLLIGLTKGDRQKIMTICTIDVHARDVAMVMLKKVDNAQAFQWLSQLRHRWADD
1801 KHCVANICDAQFKYSYELGNTPRLVITPLTDRCYITLTQSLHVMGAPAGPAGTGKTEKDLGRALGIMVYVFCSEQMDYKSCGNITKGLSQTGAW
1901 GCFDEFNRISVEVLSVVAQVQKQVDAIRDKKERFMFMEIEISLIPSVGIFITMNPYAGRTPELNLKALFRPCAMVVPDFLEICEIMLVAEGFLDARL
2001 LARKFILTLYTLCKELLKQDHYDQGLRAIKSVLVWAGSLKRGDPQRPEQVLMRALDRFNPVKIVSDPTVFMGLIGDLPALDVPRRRDMDFEYVVKQS
2101 TDLKLQAEFSVLKVVQLEELLAVRHSVFVIGNAGTSQVGLKVLNKTYSNMKRPVLDLNPKAVTNDELFGINPATREWKDGLFSVIMRDMNSITH
2201 DGPKWLDQDIDPMWIESLNTVMDNKVLTASNERILPTPSMRLLFEISHLKTATPATVSRAGILYNPDLGWNPIVTSWIDTREVQSERANLTILF
2301 DKYLPDLTLRVRFKIIPIPEQSMVQMLCYLLECLTPTNPADCPKELYELVFVFAISWAFGSMGFQDQDLYRVEFSKWIITFKTIKFPNGQTVF
2401 DYFDQESKFLPWEKVPVFLDPEIPMAQVLTNETTRVRFMDLLMERGPVMLVGNAGLSVLDGKLSNLGDSMVANVPFNYYITSEMLQRV
2501 LEKPLEKAGRNYPGPGLKLYVIFDDMMPEVDITYGVPHTLIRQHMDYKHWDROKLTKEIKHCQYVSCMNPAGTSFTNSRLQRHFCVFLSFPQ
2601 QDALSTIYNSILSHLANISVSNALQKLSPTVTSATLDLHKVAGSFLPTAIFHYVFNRLDLSNVFGQLLYSGDPLKAPIDFARLWMHCQRVYGDQKH
2701 INDDQIEAFEKLVLEYAKKFFEDVDEEAKAPNIIHCFATGIDGPKYMQVNPWELNKLVEALDYNEINAVMNLVLFEDAMQHVCRINRILESPRGM
2801 ALLVGVGGSGKSLARASYISSLEVFQILTRKGYGIPDLKDLATVCMKAGLKNIGTVFLMTDAQVDEKFLVLINDLASEGIPDLFADDEVENIIGG
2901 VRNEVKGMLQDGTRENCWKFIDRLRRLQTLVLCFSVPGTTLRVRSRKFPAVVNCTSIDVFEHPQEAQVSVSMRFLDEVELKGDIKKSIAEFMAYVHV
3001 SVNESSKLYLTNERRNYNTPKSFLEQIKLYESLLSMKSKELTAKMERLENGLTQKSTAQVDDLKAKLASQVEELAKNEDAKLQVGVTEKYSK
3101 EKATADDEKKVAIINEEVSKAKDCSDLAKAPALLAAGEALNTLNKMLTELKSFSPSAVLKVAAMVLLAPNGKIPKDRSWKAAKVMYKIDA
3201 FLDLSIYNOKENIHENCKLSIKELYNDPEFEPEYIKGKSLAAGLCSWVNVIVKFNVYCDVEPKRIALQKAMDELKAAQDKLAIKAKIAELDAMLAE
3301 TAQFEKATSDKLCQQAEEATSRITLANRLVGLLASEVVRWGEAVANFKIQEKLPGDVLITAFVSYIGCFTKTYRVDLQERMWLPFLSKQDKPIPI
3401 EGLDLSMLTDDADIAVWNEGLPSDRMSTENATILNQCQRWPLMDPQLQGIKWKQYGGDLRVIRIGQGYLDTENAISSGDTVLNMEESIDPV
3501 LDPVLGRNTIKKGRYIKIGQKEVEYNPDFRLILQTLKLANPHYKPEMAQTTLINFVTRDGLDQLLANVAGERPDLKLSDLTKQGNDFKILKELE
3601 DNLLSRLLSAEGNFGDALTALVENLETTKRTAAEISVKVEAKVEKVINARELYRPAARASLLYFILDNLKNIPYQFSLKAFNTVFLPIARAEP
3701 EDVKERVVNDICITYSVFYITRGLFEADKLITTOVAQVLLMKKEIAQNELDLRLFPQVGLTSPVDFTNSAWGAKLSLAMEDFRNLDORDIEGS
3801 AKRWKFFVESECEKEKFPQEKWKNKALQKLCMMRLADRMSYAVRNFIIEKLGSKYVEGRQVEFAKSYEETDPATPVFFILSPGVDPKDEALGKKL
3901 GFTFDNNFNHNSLGGQGEIVAEQCMDLAAKEGHVILQNIHLVAKWSTLEKKLEQYVSGSHDSYRVYSAPAGSPEAHIPQGLLESSIKITNEPPT
4001 GMFANLHKALYNFQDQTELCAREAEFKVILFALCYFHAVVCERQKFGPQGNRSYPFNTGDLTISVNVNLYNLEANSKVPWQDLRYLFGIIMYGGHITD
4101 DWRRRLCRTYLEEYMAPEMDGELYLAGFPVPNPSDYKQHYQIDEILPPESPYLYLHPNAEIGFLTDESNDLKFIVLELQPRDAGGGGGSSREEK
4201 IKSLDEIVEKLEPEEFNMMEIMAKVEDRTPVVVAQFECERMNLTSEIRSLKELDLGLKELTITPDMEDLSNALFLDQIPATVKKRAYPSFGLTSW
4301 YADLLQRIKELEQWADFALPNVVLGGFFNPQSLTAIMGSMARKNEWPLDKMLCQDVTYKKNKEDFSAPREGSVHGLFMGARWDTQTNHJADARL
4401 KELAPNMPVIFIKATPVQDKQTRNIYECVPVYTKQGRGTFVWTFNLKSKEAKAKWTLAGVALLLQV

```

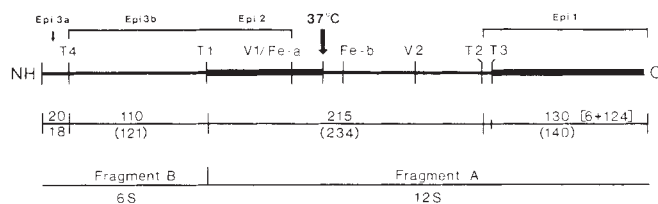


FIG. 2 Map of the principal tryptic and photolytic cleavage sites on the dynein β chain shown with the revised polarity indicated by the sequence data in Fig. 1. Numbers above the middle line give the masses of the tryptic peptides ($M_r \times 10^{-3}$) estimated from their electrophoretic migration^{11,25}. Numbers in parentheses below this line give the calculated masses assuming that no additional digestion occurs at the C termini. For consistency, the peptides are referred to by their apparent M_r s on electrophoresis. The N termini of the 110K, 215K, 130K and 124K tryptic peptides are indicated as T4, T1, T2 and T3, respectively. Thick regions of the map indicate peptides that are stable to trypsin at 37°C²⁵. The primary consequences of the revised polarity of the map are that fragment B, which sediments at 6S and possibly represents the tail structure of dynein β heavy chain seen in the electron microscope²⁷, derives from the N-terminal region of the heavy chain, and fragment A, which sediments at 12S and may represent the globular head, derives from the C-terminal region. In addition, the larger peptide resulting from V1 photocleavage (HUV1) now lies on the C-terminal side of the smaller LUV1 peptide.

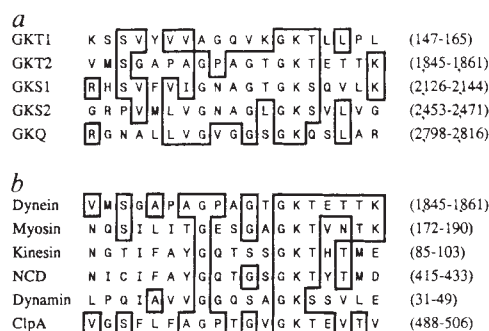


FIG. 3 **a**, Alignment of the putative nucleotide-binding sites in the dynein β heavy chain. Residues identical in two or more sites are boxed. **b**, Alignment of the probable hydrolytic ATP-binding sites of dynein β chain (GKT2), myosin²⁹, kinesin heavy chain³⁰, *ncd* protein^{31,32} (a minus end-directed kinesin-like protein), dynamin²⁸ and ATP-dependent Clp protease, A subunit¹⁹ (*Escherichia coli*). Residues identical with the dynein GKT2 sequence are boxed.

2,809 in the GKQ site a candidate for the locus of V2 cleavage. But it remains to be determined whether the AGQVKGKT sequence at 154-161 and the GVGSGGKQ sequence at 2,805-2,812 are part of functional nucleotide-binding sites, evolutionary remnants or coincidence.

As dynein ATPase activity decreases in proportion to the fraction of photocleavage at the V1 site¹⁰, it is unlikely that any of the other binding sites hydrolyses ATP at an appreciable rate under the conditions tested. On the other hand, general support for multiple nucleotide-binding sites in the β heavy chain is provided by data from photoaffinity labelling with the substrate analogues 2-azidoATP and 8-azidoATP. In both sea urchin and *Chlamydomonas*, a complex pattern of incorporation is observed with label in both peptides formed on V1 or V2 photocleavage^{22,23}, consistent with binding of the analogue at multiple sites.

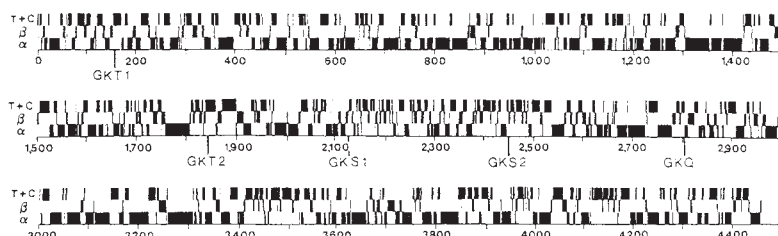
The secondary structure of the β heavy chain has been analysed by the methods of Gascuel and Golmard²⁴ (Fig. 4), Garnier, and Chou and Fasman (data not shown). Although they differ in detail, all three methods agree in predicting an α/β -type structure with about 35% coil along the whole length of the chain. The 110K peptide making up fragment B appears fairly homogeneous along its length, and although it appears to have a higher proportion of α helix between residues 703-1,025, none of the individual helices is longer than about 30 residues. The 215K peptide of fragment A shows a longitudinal differentiation that seems to be related to the putative nucleotide-binding sites. The proposed hydrolytic binding site at 1,845-1,861 is located in a region of generally low helicity, bounded on the amino side by the highly helical region at 1,760-1,813, and on the carboxy side by the region of average helicity (1,922-2,127) that separates it from the two GKS sites. A long region (2,128-2,543) of relatively low helix content encompasses the two GKS sites, and is bounded on the carboxy side by another region of average helicity. Apart from these regions located near the putative nucleotide-binding site, the 215K peptide contains two regions of α/β -type structure with well above average helicity near its N terminus (1,192-1,421) and C terminus (2,975-3,223). After a short gap around the tryptic cleavage site at 3,238, the region of high helicity continues on from the C terminus of the 215K peptide into the N-terminal region of the 130K peptide of fragment A (3,260-3,314). The rest of the 130K peptide seems

to consist of five to six alternating regions of relatively low and high helicity, with the string of glycines at 4,188 presumably representing a particularly flexible region of the chain. The above-average helicity in the N-terminal 92K region of the 215K peptide agrees with the finding that this region is relatively stable to mild heat, whereas its C-terminal region is destabilized by heat or low concentrations of methanol but can be stabilized by ATP²⁵.

In addition to this analysis, we have searched the β -chain sequence for regions containing a minimum of four consecutive repeats of the heptad hydrophobic pattern²⁶ that characterizes α -helical coiled coils like those in the myosin rod and the folded α helices like those in spectrin and dystrophin. A total of 752 residues of the β chain (17%) lie in such repeats, mostly in short clusters of four to six heptads. There are, however, two long stretches of unbroken heptad hydrophobic repeat that seem likely to fulfill some special function in the β chain. These are located at 3,028-3,147 (120 residues) near the C terminus of the 215K peptide and at 3,255-3,304 (50 residues) near the N terminus of the 130K peptide, both in regions predicted to be highly α -helical (Fig. 4). The secondary hydrophobic repeat at the fourth residue that is particularly characteristic of a coiled coil structure²⁶ appears clearly only towards the carboxy end of the second cluster. One possible interpretation is that the heptad clusters correspond to the short projection on the globular head, with the β chain looping out and then folding back on itself in a leucine zipper type structure. Alternatively, they could correspond to a region of the β -chain that interacts with a similar structure on another polypeptide, either the α heavy chain or one of the intermediate chains in the outer arm, or on the surface of the microtubule.

Use of the computer program FASTA²⁷ to search the GenBank and NBRF databases for homology to the dynein β chain sequence showed that the hydrolytic site most closely matched that of one of the two ATP-binding sites of Clp protease of *Escherichia coli*¹⁹, with 53% identity over 36 residues (Fig. 3b, and data not shown). An identity of about 18% between residues, 3,000-3,360 on the β chain and the rod domains of myosin from various species is attributable to both having a similar heptad repeat rather than a true homology. We have identified no statistically significant similarity in sequence to the putative microtubule-binding regions of MAP1B, MAP2, tau or kinesin.

FIG. 4 Secondary structure of the β heavy chain predicted by the procedure of Gascuel and Golmard²⁴. The lowest set of bars indicates regions of predicted α helix, the middle set predicted β sheet and the top set the sum of coil and β turn. The numbers represent amino-acid positions as in Fig. 1. The five putative nucleotide-binding sites are indicated. The overall percentage of α helix predicted with the standard parameters used is 48%, somewhat greater than the value of 32% predicted by circular dichroism³³.



But the relative high concentration of proline (11%) and glycine (12%) residues in a region close to the C terminus of the β chain (4,117–4,213) is reminiscent of the composition of the C-terminal 100 residues of dynamin, which may be involved in microtubule-binding²⁸.

The dynein β chain differs from the cytoskeletal motor molecules myosin and kinesin in that the head-tail domain structure seen in the electron microscope^{29,30} is not readily apparent from the predicted secondary structure. Identification of functional regions will be easier when the sequence of dynein heavy chains from other organisms reveals which regions are most conserved. □

Received 22 May; accepted 24 June 1991.

- Goodenough, U. W. & Heuser, J. A. *J. molec. Biol.* **180**, 1083–1118 (1984).
- Sale, W. S., Goodenough, U. W. & Heuser, J. E. *J. Cell Biol.* **101**, 1400–1412 (1985).
- Smith, E. F. & Sale, W. S. *Cell Motil. Cytoskel.* **18**, 258–268 (1991).
- Ogawa, K. *et al. Cell Motil. Cytoskel.* **16**, 58–67 (1990).
- Vale, R. D., Soll, D. R. & Gibbons, I. R. *Cell* **59**, 915–925 (1989).
- Sale, W. S. & Fox, L. A. *J. Cell Biol.* **107**, 1793–1798 (1988).
- Ow, R. A., Mocq, G., Tang, W.-J. Y. & Gibbons, I. R. *J. biol. Chem.* **262**, 3409–3414 (1987).
- Vale, R. D. *A. Rev. Cell Biol.* **3**, 347–378 (1987).
- Gibbons, I. R. *J. biol. Chem.* **263**, 15837–15840 (1988).
- Gibbons, I. R. *et al. J. biol. Chem.* **262**, 2780–2786 (1987).
- Mocq, G., Tang, W.-J. Y. & Gibbons, I. R. *J. Cell Biol.* **106**, 1607–1614 (1988).
- Ogawa, K. *Proc. Japan Acad.* **67B**, 27–31 (1991).
- Ogawa, K. *Proc. Int. Echinoderm Cong. Tokyo* (in the press).
- Foltz, K. Asai, D. *J. Cell Motil. Cytoskel.* **16**, 33–46 (1990).
- Garber, A. T., Retief, J. D. & Dixon, G. H. *EMBO J.* **8**, 1727–1734 (1989).
- Walker, J. E., Saraste, M., Runswick, M. J. & Gay, N. J. *EMBO J.* **1**, 945–951 (1982).
- Cremon, C. R., Long, G. T. & Grammar, J. *Biochemistry* **29**, 7982–7990 (1990).
- Cremon, C. R. *Biophys. J.* **59**, 513 (1991).
- Gottesman, S., Clark, W. P. & Maurizi, M. R. *J. biol. Chem.* **265**, 7886–7893 (1990).
- Hayashi, M. & Higashi-Fujimi, S. *Biochemistry* **11**, 2977–2982 (1972).
- Mocq, G. & Gibbons, I. R. *J. biol. Chem.* **265**, 2917–2922 (1990).
- Tang, W.-J. Y. & Gibbons, I. R. *J. biol. Chem.* **262**, 17728–17734 (1987).
- King, S. M., Haley, B. E. & Witman, G. B. *J. biol. Chem.* **264**, 10210–10218 (1989).
- Gascuel, O. & Gilmard, J. L. *CABIOS* **4**, 357–365 (1988).
- Mocq, G., Farias, J. & Gibbons, I. R. *Biochemistry* (in the press).
- McLachlan, A. D. & Karn, J. *J. molec. Biol.* **164**, 605–626 (1983).
- Pearson, W. R. & Lipman, D. J. *Proc. natn. Acad. Sci. U. S. A.* **85**, 2444–2448 (1988).
- Obar, R. A., Collins, C. A., Hammarback, J. A., Shpetner, H. S. & Vallee, R. B. *Nature* **347**, 256–261 (1990).
- Warrick, H. M. & Spudis, J. A. *A. Rev. Cell Biol.* **3**, 379–421 (1987).
- Yang, J. T., Layman, R. A. & Goldstein, L. S. *B. Cell* **56**, 879–889 (1989).
- Endow, S. A., Henikoff, S. & Soler-Niedziela, L. *Nature* **345**, 81–83 (1990).
- McDonald, H. B. & Goldstein, L. S. *B. Cell* **61**, 991–1000 (1990).
- Mocq, G. & Gibbons, I. R. *Biochemistry* **29**, 48–39–4843 (1990).
- Gibbons, I. R. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).

ACKNOWLEDGEMENTS. We thank K. Ogawa for kindly sending us copies of refs 12 and 13 before publication. W. J.-Y. Tang for double-strand sequencing, N. S. Ching, G. J. Dolecki, P. J. Harwood, C. A. Phillipson and H. Ren for assistance, and N. Reimer for microsequencing the tryptic peptides. This work has been supported in part by the NIH to I.R.G. and B.H.G., and the NSF to D.J.A. The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the discussion number X59603.

Four ATP-binding sites in the midregion of the β heavy chain of dynein

Kazuo Ogawa

Department of Cell Biology, National Institute for Basic Biology, Okazaki 444, Japan

THE 'motor' proteins of eukaryotic cells contain specialized domains that hydrolyse ATP to produce force and movement along a cytoskeletal polymer (actin in the case of the myosin family; microtubules in the case of the kinesin family and dyneins). There are motor-protein superfamilies in which each member has a conserved force-generating domain joined to a different 'tail' which conveys specific attachment properties (see ref. 1 for a review). The minus-end-directed microtubule motors, the dyneins², may also constitute a superfamily of force-generating proteins with distinct attachment domains³. Axonemal outer-arm dynein from sea urchin spermatozoa is a multimeric protein consisting of two heavy chains (α and β) with ATPase activity, three intermediate

chains and several light chains⁴. Here I report the sequence of cloned complementary DNA encoding the β heavy chain of a dynein motor molecule. The predicted amino-acid sequence reveals four ATP-binding consensus sequences in the central domain. The dynein β heavy chain is thought to associate transiently with a microtubule during ATP hydrolysis⁵, but the ATP-dependent microtubule-binding sequence common to the kinesin superfamily is not found in the dynein β heavy chain. These unique features distinguish the dynein β heavy chain from other motor protein superfamilies and may be characteristic of the dynein superfamily.

Immunological tests show that sea urchin embryo cilia and sperm flagella have different dynein α heavy chains, but similar β heavy chains (D β HC)⁶, indicating that D β HC is a highly conserved protein. I isolated D β HC complementary DNA by screening an unfertilized egg poly(A)⁺ RNA expression library with antibody raised against axonemal dynein heavy chains. The nucleotide sequence and the deduced amino-acid sequence of the constructed full-size cDNA are shown in Fig. 1. The nucleotide sequence contains a single long reading frame capable of encoding a protein of 4,466 amino-acid residues with an unmodified relative molecular mass of 512,000 (M_r , 512K).

The secondary structure of the D β HC amino-acid sequence was analysed according to ref. 7 and 8 (Fig. 2). Although such methods are of limited accuracy in predicting the conformation at a given residue or for short regions, they were useful in this case for predicting the characteristics of large regions of the molecule. There are two long α -helical dominant regions (termed $\alpha 1$ and $\alpha 2$) in the sequence, which suggests that the D β HC is composed of three large β -structure dominant domains (termed the N, M and C domains) separated by these regions. The M domain is split by short α -helical dominant regions into four smaller β -structure dominant regions, and the C domain is similarly split into three smaller regions. Although secondary structure analysis predicts that the $\alpha 1$ region should be rich in α helix, there are no long uninterrupted heptapeptide repeats such as are found in the α -helical coiled-coil regions of filamentous proteins like myosin, tropomyosin and spectrin. The $\alpha 2$ region contains two leucine heptapeptide repeats predicted to be largely α -helical at residues 3,028–3,083 and 3,262–3,303, with interruption; these could favour generation of a large globular structure from the M and C domains.

Tryptic digestion of D β HC gives rise to an ATPase-containing fragment (fragment A) composed of f2 (M_r , 190K) and f3 (135K) peptides, and an A-subfibre microtubule-binding fragment B composed of f1 peptide (130K)^{9,10}. These three peptides are arranged in the order f1, f2 and f3 from the amino to the carboxy terminus of the D β HC molecule¹¹. Direct sequencing of the f2 and f3 peptides identified trypsin cleavage sites at Lys 1,191 of the $\alpha 1$ region (termed the T1 site) and at Arg 3,323 of the $\alpha 2$ region. Therefore, the N, M and C domains correspond to the β -structure dominant regions of peptides f1, f2 and f3, respectively. The f2 and f3 peptides remain together in fragment A after trypsin digestion, whereas f1 is detached, in agreement with the possibility that the α -helical coiled-coil regions of the $\alpha 2$ region might promote the association of the M and C domains in a larger globular structure. Quick-freeze deep-etch electron microscopy of D β HC¹² shows that it has a pear-shaped head, which may correspond to fragment A, M and C domains, and an irregularly shaped stem, which may correspond to the fragment B, N domain.

In view of the ATP-hydrolysing capacity of D β HC, I inspected the predicted amino-acid sequence for features in common with other nucleotide-binding proteins. The consensus sequence (in one-letter code) indicative of ATP binding is GXXXXGKT(or S), where X is any amino acid¹³. This was found at three positions in the D β HC sequence, beginning at Gly 1,852 (termed the GKT site), Gly 2,133 (GKS1 site), and Gly 2,460 (GKS2 site). GXXXXSGK is an ATP-binding sequence of adenylate kinase¹³ and is also found in the D β HC sequence, beginning at Gly 2,805 (SGK site). Therefore, molecular cloning

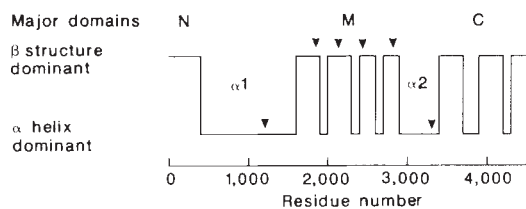


FIG. 2 Predicted secondary structure for the D β HC obtained by methods described in refs 7 and 8. These methods distinguish β -structure dominant regions and α -helix dominant regions, which are pictured here in relation to the numbered positions of the amino-acid residues deduced from the D β HC sequence. The four arrowheads on the M domain indicate the GKT, GKS1, GKS2 and SGK sites of the ATP-binding consensus sequences from the left. The two arrowheads on two long α -helix dominant regions define the trypsin cleavage sites.

of D β HC cDNA reveals the presence of four putative ATP-binding sites in the midregion of the molecule. Each of these sites may be located in one of the four predicted β -structure dominant regions of the M domain.

Does this multiplicity of putative ATP-binding sites indicate that D β HC possesses several active sites for ATP hydrolysis? Ultraviolet irradiation of D β HC in the presence of magnesium ion, ATP and vanadate cleaves the D β HC at a single site, the V1 site¹⁴. Because vanadate can potentially inhibit dynein ATPase, probably as a result of occupying a site normally reserved for the γ -phosphate of ATP, the V1 site probably lies in the hydrolytic domain of the D β HC. As the V1 site is located 70K to the carboxy terminal side of the T1 site¹⁴, it corresponds to the GKT site. The D β HC can be covalently modified with the hydrolysable photoaffinity ATP analogue, 8-azido adenosine 5'-triphosphate (8-N₃ATP)¹⁵. As 8-N₃ATP is a competitive inhibitor of dynein ATPase from *Chlamydomonas*¹⁶, the modified portion should correspond to an ATP-hydrolysis site of D β HC. This site is located 170K to the carboxy-terminal side of the T1 site¹⁵, suggesting that the SGK site may correspond to this 8-N₃ATP-binding site. Therefore, both the GKT and SGK sites may be functional ATP hydrolysis sites. The joint method⁷ for predicting the secondary structure of D β HC makes all four ATP-binding sites equivalent, suggesting that the two GKS sites may also be functional ATP hydrolysis sites. Motor proteins such as myosin and kinesin have one ATP-binding consensus sequence in the molecule, so the presence of several ATP-binding sites, perhaps as active sites in the motor domain, may be a characteristic of the dynein superfamily.

The outer arm dynein contains two types of microtubule-binding site. The arms are thought to associate transiently with the B-subfibrils of the adjacent axonemal doublet microtubules during the ATP-hydrolytic cycle, and also to have permanent attachments on the A-subfibrils. The N domain may contain the microtubule-binding site for attachment to the A-subfibre. MAP2 (ref. 17) and tau¹⁸ proteins form stable complexes with microtubules; they share a region of homologous sequences at their carboxy termini, which contains three 18-residue repeats and which has been proposed as the microtubule-binding site. These repeated sequences are not found in the N domain of the D β HC sequence. Members of the kinesin superfamily of microtubule motor proteins also associate transiently with microtubules during the ATP-hydrolytic cycle. They share a region of sequence homology extending from the ATP-binding site towards the carboxy terminal end of the molecule, which has been suggested to constitute the ATP-dependent microtubule-binding site¹⁹. This sequence homology was not found in the C domain or anywhere else in the D β HC sequence. D β HC seems to be a member of a new family of microtubule-binding motor proteins with unique microtubule-binding sequences which are unlike those of MAP2, tau, or members of the kinesin superfamily. □

Received 22 May; accepted 24 June 1991.

1. Vale, R. D. & Goldstein, L. S. B. *Cell* **60**, 883-885 (1990).
2. Gibbons, I. R. *J. Cell Biol.* **91**, 107s-124s (1981).
3. Ogawa, K. *et al. Eur. J. Cell Biol.* **43**, 3-9 (1987).
4. Tang, W.-J. Y. *et al. J. Biol. Chem.* **257**, 508-515 (1982).
5. Sale, W. S. & Fox, L. A. *J. Cell Biol.* **107**, 1793-1797 (1988).
6. Ogawa, K. *et al. Cell Motil. Cytoskel.* **16**, 58-67 (1990).
7. Nishikawa, K. & Noguchi, T. *Meth. Enzym.* **202**, 31-44 (1991).
8. Nishikawa, K. & Ooi, T. *J. Biochem.* **100**, 1043-1046 (1986).
9. Ogawa, K. & Mohri, H. *J. Biol. Chem.* **250**, 6476-6483 (1975).
10. Ow, R. A. *et al. J. Biol. Chem.* **262**, 3409-3414 (1987).
11. Ogawa, K. *Proc. Japan Acad.* **67B**, 27-31 (1991).
12. Sale, W. S. *et al. J. Cell Biol.* **101**, 1400-1412 (1985).
13. Walker, J. E. *et al. EMBO J.* **1**, 945-951 (1982).
14. Gibbons, I. R. *et al. J. Biol. Chem.* **262**, 2780-2786 (1987).
15. Tang, W.-J. Y. & Gibbons, I. R. *J. Biol. Chem.* **262**, 17728-17734 (1987).
16. Pfister, K. K. *et al. J. Biol. Chem.* **259**, 8499-8504 (1984).
17. Lewis, S. L. *et al. Science* **242**, 936-939 (1988).
18. Lee, G. *et al. Science* **239**, 285-288 (1988).
19. Yang, J. T. *et al. Cell* **56**, 879-889 (1989).
20. Kozak, M. *Cell* **44**, 283-292 (1986).
21. Matsudaira, P. *J. Biol. Chem.* **262**, 10035-10038 (1987).

ACKNOWLEDGEMENTS. I thank K. Nishikawa of the Protein Engineering Research Institute for predicting the secondary structure of the D β HC; C. J. Brokaw of the California Institute of Technology for critically reading the manuscript; H. Mori and M. Hayashi for synthesising oligonucleotides; H. Kajiura for amino-acid sequencing; and K. Inaba for preparing f3 peptide from sperm 21S dynein. The sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence Databases under the accession number D01021.

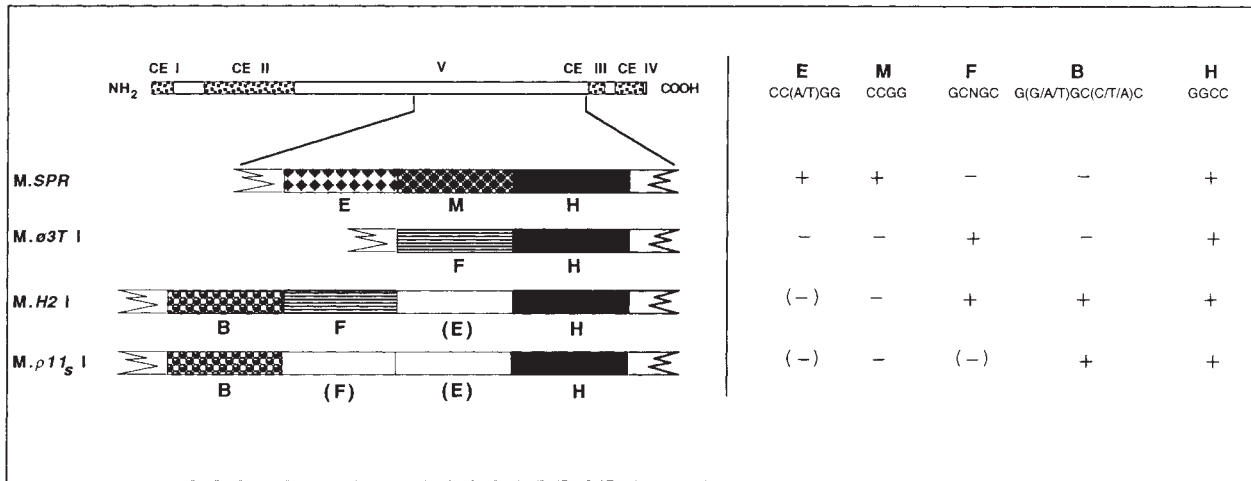
'Pseudo' domains in phage-encoded DNA methyltransferases

C. Lange, A. Jugel, J. Walter, M. Noyer-Weidner & T. A. Trautner

Max-Planck-Institut für molekulare Genetik, Ihnestrasse 73, 1000 Berlin 33, Germany

5-Cytosine-DNA-methyltransferases, which are found in many organisms ranging from bacteriophages to mammals, transfer a methyl group from S-adenosylmethionine to the carbon-5 of a cytosine residue in specific DNA target sequences¹. Some phage-encoded methyltransferases methylate more than one sequence: these enzymes contain several independent target-recognizing domains each responsible for recognizing a different site. The amino-acid sequences of these multispecific methyltransferases reveal that some enzymes in addition carry domains that do not contribute to the enzymes' methylation potential, but strongly resemble previously identified target-recognizing domains. Here we show that introducing defined amino-acid alterations into these inactive domains endows these enzymes with additional methylation specificities. Gel retardation analysis demonstrates that these novel methylation specificities correlate with the acquisition of additional DNA-binding potential of the proteins.

Comparisons of the primary structures of about 40 different methyltransferases revealed at least four highly conserved amino-acid sequence motifs in identical order but variable in spacing (Fig. 1). At the same relative location with respect to a core of conserved sequences, all methyltransferases have in addition an extended region that is variable in size and composition between different enzymes (Fig. 1). The conserved core sequences are thought to be responsible for common functions, such as S-adenosylmethionine (SAM) binding and the methyl group transfer²⁻⁵, whereas target recognition has been assigned to the variable region. Enzymes with multiple specificities have several independent domains within the variable region of the enzymes, each of which recognizes a different DNA target⁵⁻⁸. These target-recognizing domains (TRDs) represent contiguous, consecutively arranged amino-acid segments, comprising from 38 to at least 59 amino acids (T.A.T. *et al.*, unpublished results). Although TRDs specifying the recognition of different targets vary in size and amino-acid composition, they share several conserved amino acids, providing a 'consensus' sequence (Fig. 2)².



H2 (ref. 9), showing the relative positions of the different TRDs (refs 5, 8), are shown below. E, M, F, B, and H identify TRDs recognizing CC(✓)yGG (*EcoRII*), CCGG (*MspI*), GCNGC (*Inu4HT*), G(G/A/T)GC(C/T/A)C (*Bsp1286*) and GGCC (*HaeIII*) target sequences, respectively. Inactive TRDs are indicated by bracketed letters. The methylation capacities of the different phage methylases (*M.SPR*, *M.φ3T1*, *M.H2I* and *M.φ11sI*) are shown in the right-hand part of the figure. (—). Activatable specificities.

tion by these enzymes does not render DNA resistant to degradation by the relevant cognate restriction endonucleases. Neither do such enzymes bind to oligonucleotides containing the targets for either *EcoRII* or *Fnu4HI* (Fig. 4) recognized by TRDs E or F. The 'inert' TRDs do not have any undetected specificity because the partially purified enzymes do not incorporate methyl groups into DNA multiply methylated *in vivo* at *HaeIII*, *Bsp1286*, *Fnu4HI* and *EcoRII* target sites (data not shown).

[illegible]

NATURE • VOL 352 • 15 AUGUST 1991

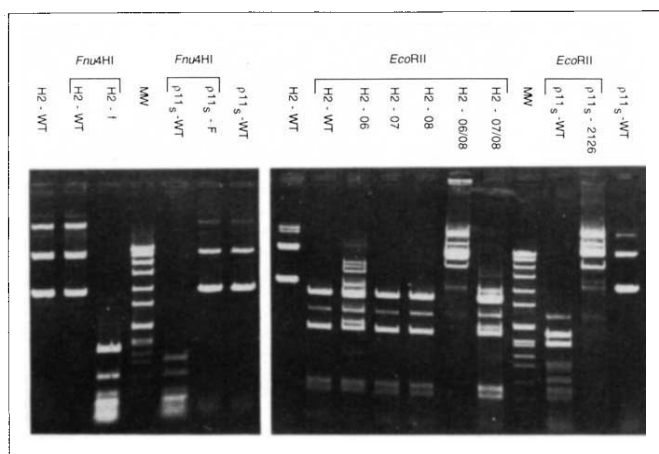


FIG. 3 Electrophoretic separation of plasmid DNA modified by *M.H2I*, *M.ρ11sI* and their mutant derivatives after incubation with restriction endonucleases *R.Fnu4HI* and *R.EcoRII*. Methyltransferase genes with original (WT) and mutagenized TRDs were cloned into plasmid pMS119, which is a derivative of pJF118¹⁴, and transformed into *E. coli* K12 strain GM271 (*mcrA*⁻, *mcrB*-1, *hsdR*-2, *dcm*-6; obtained from M. G. Marinus). Following preparation of plasmid DNA from these transformants¹⁵, DNA (1 µg) was incubated for 1 h with an excess of the appropriate restriction enzymes and subsequently loaded on a 1% agarose gel. The proteins that were expressed by the different plasmids are indicated at the top of the figure as described in Fig. 2. SPP1 DNA digested with *R.EcoRI* served as a size marker (MW)¹⁶.

function of this TRD, as this is the only position where the sequences of *M.H2I* and *M.ρ11sI* diverge. Indeed, changing Glu 291 to Gly provided *M.ρ11sI* with an additional *Fnu4HI* specific methylation capacity, whereas the reverse mutation introduced into *M.H2I* eliminated methylation of *Fnu4HI* sites. Apparently other amino-acid differences between *M.φ3TI* and *M.H2I* are irrelevant to the function of this domain.

The inactive E-like TRDs from *M.ρ11sI* and *M.H2I* differ from the *M.SPR* sequence at seven amino-acid positions (Fig.

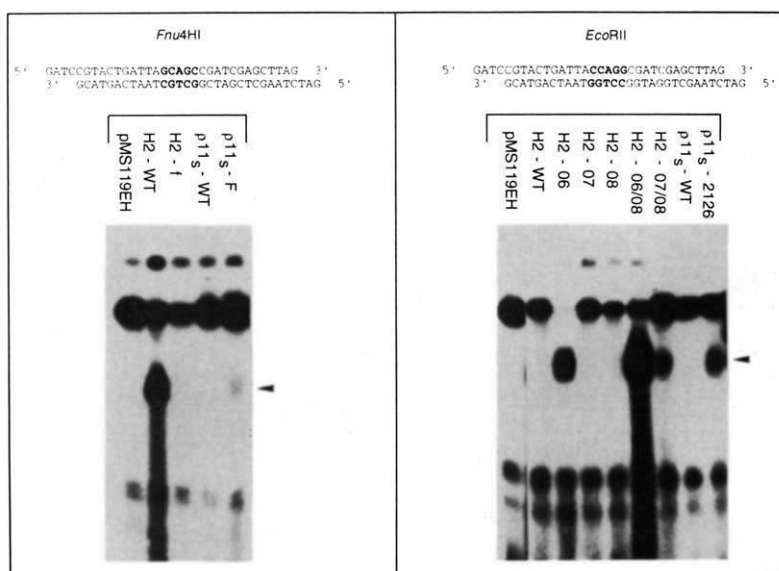
2b). Our attempts to activate the H2-encoded domain concentrated on those amino acids found in the TRD consensus sequence. According to this rationale we introduced the following single amino-acid exchanges into *M.H2I*: Glu 311 → Asn, Ser 325 → Pro and Ser 338 → Ala (Fig. 2b). The conversion from Ser 325 into Pro induced an *EcoRII* specific methylation activity of the enzyme (H2-06 in Fig. 2b), which makes the plasmids expressing the mutagenized gene partially resistant to *EcoRII* specific restriction (Fig. 3). Introduction of Asn 311 (H2-07) and Ala 338 (H2-08) had no effect on the phenotype of *M.H2I*. In addition, double mutants with changes Ser 325 → Pro/Ser 338 → Ala (H2-06/08) and Glu 311 → Asn/Ser 338 → Ala (H2-07/08), gave full or partial resistance to *EcoRII* restriction (Fig. 3). These results also provided some rationale for the mutagenesis of the inactive domain (E) of *M.ρ11sI*, which differs from the corresponding *M.H2I* sequence at three positions. Based on this and the TRD consensus sequence, we replaced Asp 315 of *M.ρ11sI* with Gly present at this position in the active E TRD of *M.SPR* (Fig. 2b), resulting in *EcoRII* specific methylation activity in *M.ρ11sI* (ρ11s-2126 in Figs 2b and 3). By gel retardation we could show that the amino-acid exchanges responsible for activating the nonfunctional TRDs also provide them with the expected new DNA binding specificities (Fig. 4).

These results support the concept of a strictly modular organization of multispecific methyltransferases. In these enzymes a highly conserved core provides the DNA methylation function, and this can be associated with various DNA recognition domains, which determine the methylation specificity. These TRDs function irrespective of their relative location to each other. Thus, the active E TRD of *M.SPR* and the activatable related domains of *M.H2I* or *M.ρ11sI* occupy different relative locations within the variable region of these enzymes (Fig. 1). Similarly, the activated F domain of *M.ρ11sI* has a location in this methyltransferase that is different from the equivalent domain in *M.φ3TI*.

The inert TRDs are probably 'down' mutations of active domains. This follows from the observation that the number of amino acids in inert and the matching active TRDs is the same (the TRDs identified in multispecific methyltransferases are characterized by distinct numbers of amino acids), and by their

FIG. 4 Gel retardation analysis performed with *M.H2I* and *M.ρ11sI* and mutant derivatives. Crude extracts of *E. coli* K12 cells overproducing the various methyltransferases indicated above the gel tracks were incubated with the synthetic oligonucleotides shown, containing either a *Fnu4HI* or a *EcoRII* target site. A crude extract prepared from cells harbouring pMS119EH, the vector used for methyltransferase overproduction, served as control. DNA-protein complexes were separated by electrophoresis on nondenaturing polyacrylamide gels. In addition to a 'background' complex, detectable also with the control extract, specific methyltransferase-DNA complexes, whose position is indicated by the arrow, were formed in some of the extracts analysed. Binding of the *Fnu4HI*-specific oligonucleotide was only detectable with methyltransferases methylating *Fnu4HI* target sites (H2-WT, ρ11s-F). The same is true for the *EcoRII*-specific oligonucleotide, where the formation of specific complexes is observed only with those enzymes which contain partially (H2-06, H2-07/08) or fully active (H2-06/08, ρ11s-2126; see Fig. 2) *EcoRII* specific TRDs.

METHODS. Cultures (2 ml) of *E. coli* K12 GM271 (for genotype see legend of Fig. 3) harbouring the different methyltransferase-expressing plasmids were induced with IPTG (2 mM) at early log phase and grown for further 4 h. The cells were collected, suspended in 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 100 mM EDTA, 5 mM DTE and disrupted by sonification, as described elsewhere (A.J. *et al.*; manuscript in preparation). Binding reactions were performed by incubation of 1 nM of the ³²P-end-labelled oligonucleotides with 2 µl of the crude extracts in buffer containing final concentrations of 50 mM Tris-HCl (pH 7.5), 50 mM NaCl, 10 mM EDTA, 5 mM



DTT and 50 µM S-adenosylhomocysteine. Samples were incubated for 5 min at room temperature followed by separation of protein-DNA complexes on 5% polyacrylamide gels at 5 V cm⁻¹.

great sequence similarity. The identity of N-terminal amino acids of active and inactive TRDs F of *M.H2I* and *M.p11sI* suggests that the inactive TRD of *M.p11sI* derived from *M.H2I*. The inactivating mutation must have been a recent event, as the nucleotide sequences of the two regions differ only in the base causing the Glu to Gly change at position 291. The nucleotide sequences encoding the inactive TRDs (E) of *M.p11sI* and *M.H2I* differ from each other in five positions, three of which account for the amino-acid differences and two of which are null. Comparing the inactive E TRDs of *M.p11sI* and *M.H2I* to the active E TRD of *M.SPR* (Fig. 2b) shows that the evolution of the inactive TRDs from an active ancestor must have followed separate avenues.

These inert TRDs represent an intriguing case of preservation of nonfunctional genetic information. This information has been preserved in the absence of positive selective pressure, which in the case of the phage-encoded modification methyltransferases would be exerted by a bacterial restriction and modification system. Negative selection, counteracting the maintenance of genetic information encoding inactive TRDs in the phage genome, has also apparently not acted.

Thus the silent TRDs represent a reserve of genetic material of general evolutionary importance as suggested, for example, by the presence of 'pseudogenes' in some bacterial and eukaryotic genomes. The inactive TRDs described here—although 'expressed' within the context of the active methyltransferases—would operationally fulfil the role of pseudogenes. For this reason we term the inactive domains 'pseudo' TRDs. Considering the modular structure of the methyltransferases discussed here it is conceivable that maintenance of these pseudo domains is significant for the generation of new specificities. □

Received 8 May; accepted 24 June 1991.

1. Razin, A., Cedar, H. & Riggs, A. D., eds. *DNA Methylation: Biochemistry and Biological Significance*, Springer Verlag, Berlin, 1984.
2. Lauster, R., Trautner, T. A. & Noyer-Weidner, M. *J. molec. Biol.* **206**, 305–312 (1989).
3. Pósfai, J., Bhagwat, A. S., Pósfai, G. & Roberts, R. J. *Nucleic Acids Res.* **17**, 2421–2435 (1989).
4. Ingrosso, D., Fowler, A. V., Bleibaum, J. & Clarke, S. *J. biol. Chem.* **264**, 20131–20139 (1989).
5. Wilke, K. et al. *EMBO J.* **7**, 2601–2609 (1988).
6. Behrens, B. et al. *EMBO J.* **6**, 1137–1142 (1987).
7. Balganesch, T. S. et al. *EMBO J.* **6**, 3543–3549 (1987).
8. Trautner, T. A., Balganesch, T. S. & Pawlek, B. *Nucleic Acids Res.* **16**, 6649–6658 (1988).
9. Lange, C., Noyer-Weidner, M., Trautner, T. A., Weiner, M. & Zahler, S. A. *Gene* **100**, 213–218 (1991).
10. Buhk, H.-J. et al. *Gene* **29**, 51–61 (1984).
11. Tran-Betoke, A., Behrens, B., Noyer-Weidner, M. & Trautner, T. A. *Gene* **42**, 89–96 (1986).
12. Kunkel, T. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 488–492 (1985).
13. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463–5467 (1977).
14. Fürste, J. P. et al. *Gene* **48**, 119–129 (1986).
15. Sambrook, J., Fritsch, E. F. & Maniatis, T. (eds) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1989).
16. Ratcliff, S. W. *Mol. gen. Genet.* **168**, 165–172 (1979).

CORRECTION

Identification of a G1-type cyclin *puc1*⁺ in the fission yeast *Schizosaccharomyces pombe*

Susan L. Forsburg & Paul Nurse
Nature **351**, 245–248 (1991)

WE have discovered an error in the DNA sequence of our *puc1*⁺ gene in Fig. 1 of the above paper. The CG pair at nucleotides 710–711 should be a GC pair. The amino acids underneath should thus be Leu Gln instead of Phe Glu. However, the protein sequence in Fig. 2 is correct. □

nature
is available in
microform.



University Microfilms

International reproduces this publication in microform: microfiche and 16mm or 35mm film. For information about this publication or any of the more than 13,000 titles we offer, complete and mail the coupon to: University Microfilms International, 300 N. Zeeb Road, Ann Arbor, MI 48106. Call us toll-free for an immediate response: 800-521-3044. Or call collect in Michigan, Alaska and Hawaii: 313-761-4700.

☐ Please send information about these titles:

Name _____

Company/Institution _____

Address _____

City _____

State _____ Zip _____

Phone () _____

University
Microfilms
International